Virginia L. Ferris. Beyond "Showing What We Have": Exploring Linked Data for Archival Description. A Master's Paper for the M.S. in L.S degree. July, 2014. 76 pages. Advisor: Timothy Shearer

Research explores current professional attitudes, applications, and potential impact of linked data in archival repositories, through eight structured interviews with archivists and early adopters of linked data and semantic web practices within the archival community. Interviews evaluate current archival practitioners' perceptions of the benefits and challenges of applying linked data practices to archival description, to examine the current barriers and potential solutions in integrating linked data with archival description. Analysis investigates ways in which linked data and the semantic web may impact archival descriptive practices and user experience in accessing archival materials, and ways to facilitate greater engagement with linked data within the archival community.

Headings:

Archives -- Cataloging

Authority control

Linked open data

Metadata

Semantic web

BEYOND "SHOWING WHAT WE HAVE": EXPLORING LINKED DATA FOR ARCHIVAL DESCRIPTION

by
Virginia L. Ferris

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Library Science.

Chapel Hill, North Carolina

July 2014

Approved by

_____

Timothy Shearer

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to thank my advisor Tim Shearer, who provided invaluable guidance from the beginning stages of this paper, with the greatest thought and care throughout the process. Jacqueline Dean and the staff of Technical Services for Manuscripts at UNC Chapel Hill's Wilson Library provided me with the best possible education in archival description, both in and out of the classroom.

I am indebted to each of the interviewees who gave so generously of their time and thoughts, and without whom this paper would not be possible.

Many thanks go to Foster's Market in Chapel Hill and Joni Mitchell, for providing me with the fuel, space, and inspiration to get things done.

Finally I am grateful to my parents, for always supporting and inspiring me to be curious and to explore the road less traveled.

INTRODUCTION

The archival finding aid seeks to accurately and meaningfully represent the physical and intellectual content of an archival collection. Archivists are involved in the work of processing collections of unpublished manuscripts materials, arranging and preserving them, and describing them so that researchers can discover and use them. As archivists handle the materials firsthand, relationships become apparent within the contents, and materials within and across collections make sense through their connections with each other. A single letter gains meaning and research value with a greater understanding of what precedes and follows it, who created it, who received it, who has preserved it over the years, and other contextual information that is preserved by using archival principles of provenance, original order, and *respect des fonds*.[1] In order to represent contextual relationships to users online, archivists rely on these traditional principles as well as on newer descriptive standards and practices, such as Describing Archives: A Content Standard (DACS), Encoded Archival Description (EAD), and Encoded Archival Context: Corporate Bodies, Persons, and Families (EAC-CPF) that seek to preserve this context for diverse users across diverse settings.[2]

---

[1] Laura Millar, *Archives: Principles and Practices* (New York, NY: Neal-Schuman Publishers, Inc., 2010).

[2] Society of American Archivists, *Describing Archives: A Content Standard* (Chicago, IL: Society of American Archivists, 2013); Jesse Brown, "More Than Just a Name: Archival Authority Control, Creator Description, and the Development of Encoded Archival Context (EAC)" (master's paper, School of Information and Library Science, University of North Carolina at Chapel Hill, 2006).

Through working firsthand with the original materials, an archivist appreciates the challenge and urgency of providing the richest description of these materials possible, and ensuring that this description be made accessible as broadly as possible. At a recent annual meeting for the Society of American Archivists, attendees participated in conversations about early experimental projects using linked data to expose archival description to broader audiences, centering on the early development of a guidebook recently published through the Tufts University Linked Archival Metadata (LiAM) Project.[3] Linked data, introduced and defined by Tim Berners-Lee as part of his vision of the "semantic web," presents new opportunities for linking information resources online as a model for opening and exchanging data using Universal Resource Indicators (URIs) that unify description and expose connections between information resources in the semantic web, building on a vision of the Internet as a web of data rather than a web of documents.[4] Linked data is most frequently expressed through Resource Description Framework (RDF) triples, using a predicate structure to express relationships between resources. According to Allemang and Hendler (2008) "the main idea of [linked data] is to support a distributed web at the level of the data rather than at the level of presentation [documents]. Instead of having one webpage point to another, one data item can point to

[3] LiAM Project: Linked Archival Metadata, "Prospectus for Linked Archival Metadata: A Guidebook," *LiAM: Linked Archival Metadata Blog*, 1 March 2013, accessed 1 July 2014, http://sites.tufts.edu/liam/deliverables/prospectus-for-linked-archival-metadata-a-guidebook/; Eric Lease Morgan and the LiAM: Linked Archival Metadata Project, *Linked Archival Metadata: A Guidebook* (version 0.99, 23 April 2014), accessed 1 July 2014, http://infomotions.com/sandbox/liam/tmp/guidebook.pdf.

[4] Tim Berners-Lee, "The Semantic Web," in *Scientific American*, 284, (2001).

another, using global references called URIs."[5] Archivists and librarians have begun to investigate the potential for linked data practices to leverage existing description in their repositories, in order to improve resource discovery and use locally and globally.

One example of this movement toward linked data can be seen in Europeana, a collective of cultural heritage collections in Europe that has created the Europeana Data Model (EDM), using semantic web and linked data practices to publish library and archival metadata.[6] The Europeana Data Model has provided a foundation for the Digital Public Library of America and other metadata harvesting services to represent complex objects and relationships within libraries and archives, and to share this data with other institutions. Robert H.P. Engels and Jon Roar Tonneson explored the use of semantic web technologies in a Digital Music Archive for collections of recordings in the Norwegian National Broadcaster (NRK) archives. Their case study found that Semantic Web techniques promoted interoperability with other archival systems, produced more robust metadata description, and greatly facilitated access and discovery of "hidden treasures" in the collections.[7]

Archives and libraries are exploring the uses of linked data in various pilot projects, including Linked Open Copac and Archives Hub (LOCAH) founded in the

---

[5] D. Allemang and J. Hendler, *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*. Boston, Massachusetts: Morgan Kaufman Publishers, 2008.

[6] Antoine Isaac, "Case Study: Enriching and Sharing Cultural Heritage Data in Europeana," *World Wide Web Consortium: Semantic Web Use Cases and Studies*, June 2012, accessed 1 July 2014, http://www.w3.org/2001/sw/sweo/public/UseCases/Europeana/.

[7] Robert H.P. Engels and Jon Roar Tonneson, "Case Study: A Digital Music Archive (DMA) for the Norwegian National Broadcaster (NRK) Using Semantic Web Techniques," *World Wide Web Consortium: Semantic Web Use Cases and Studies* (2007), accessed 1 July 2014, http://www.w3.org/2001/sw/sweo/public/UseCases/NRK/.

United Kingdom in 2010; Linked Open Data in Libraries Archives and Museums (LODLAM), a network of professionals founded in 2010; and the Tufts University Linked Archival Metadata (LiAM) Project. Adrian Stevenson of the LOCAH Project has asserted that linked data creates "links with other data sources, and allows the development of new channels into the data. Researchers are more likely to discover sources that may materially affect their research outcomes, and the 'hidden' collections of archives and special collections are more likely to be exposed and used."[8]

There remains a need to build case studies and evidence-based literature related to the implementation of linked data in archival description. This study seeks to investigate current projects underway at various archives that are experimenting with linked data, and to examine contemporary archivists' views toward this growing practice. Research methods follow the model presented by Jesse Brown (2006) in his study of the impact of Encoded Archival Context (EAC) as a newly introduced descriptive tool in the archival community. This research framework will guide an evaluation of the professional attitudes and perceptions of linked data among practitioners familiar with semantic web technologies and archival description. The study will utilize thick description to record and analyze practitioners' observations and experiences with linked data in the archival community. The study will build on existing literature and contribute original research in an effort to encourage further exploration of linked open data in archival practice.

Research Questions

[8] Adrian Stevenson, "LOCAH Project: Wider Benefits to Sector and Achievements for Host Institution," *LOCAH Project* blog, 2010, accessed 1 July 2014, http://archiveshub.ac.uk/locah/2010/08/locah-project-wider-benefits-to-sector-achievements-for-host-institution/.

The following research questions guided the study and informed structured interviews with research participants as part of original data collection:

1. How do archivists currently understand, define, and use linked data and semantic web practices?
2. What is the perceived value of linked data in archival description, according to current practitioners?
3. What are the perceived barriers to implementing linked data in archival description, according to current practitioners?
4. What may remove these barriers?
5. How might linked data impact the future of archival description?

LITERATURE REVIEW

In the absence of major empirical studies on the use of linked data in archives, case studies and institutional white papers on linked data in a variety of settings have provided the basis for a literature review in this study. By linking data within and across different collections and domains of cultural heritage on the web, a growing number of archivists and librarians have reported that linked data provides a powerful tool for adding value to their resources and enriching discovery.[9] However, within the library and archives community, obstacles have blocked the widespread implementation of linked data. Getaneh Alemu, Brett Stevens, Penny Ross, and Jane Chandler identified the benefit of linked data as a data model that facilitates "the re-usability, cross-linking, integration and sharing of data" for libraries.[10] However, their research found that barriers to its adoption include the "document-centric metadata structure" of machine-readable cataloging (MARC) used widely in libraries, a general "terminological disparity"

---

[9] Gillian Byrne and Lisa Goddard, "The Strongest Link: Libraries and Linked Data," *DLib Magazine*, Vol. 16, No. 11-12 (2012), accessed 1 July 2014, http://www.dlib.org/dlib/november10/byrne/11byrne.html; Johan Oomen, Lotte-Belice Baltussen, and Marieke Van Erp, "Sharing Cultural Heritage the Linked Open Data Way: Why You Should Sign Up" (paper presented at *Museums and the Web*, San Diego, CA, 2012), accessed 1 July 2014, http://www.museumsandtheweb.com/mw2012/papers/sharing_cultural_heritage_the_linked_open_data; Bethan Ruddock, "Linked Data and the LOCAH Project." *Business Information Review*, Vol. 28, No. 2 (2011): 105-111.

[10] Getaneh Alemu, et al, "Linked Data for Libraries: Benefits of a Conceptual Shift from Library-Specific Record Structures to RDF-based Data Models," in *New Library World*, Vol. 113, No. 11-12 (2012), 549.

between library and web-based metadata standards, and most importantly the relative complexity of linked data technologies such as RDF/XML markup, SPARQL search protocol and query language, and the OWL web ontology language.[11]

Jane Greenberg and Eva Mendez found that what they have termed Linked Open Vocabularies (LOV) may remove many of these barriers, by facilitating a "semantic web environment where vocabularies are explicitly created for the web," and therein promote interoperability and integration of data across many different vocabularies.[12] In addition to this growing understanding of the potential advantages of using linked data in libraries and archives, an increasing number of prominent international cultural heritage institutions have adopted linked data practices.[13] Examples of these institutions include the Library of Congress, Europeana, the Swedish Union catalog, the German National Library, the British National Bibliography, and the Digital Public Library of America (DPLA). Mitchell found that Europeana, the DPLA, and the Library of Congress BIBFRAME initiative all demonstrated similar high-level frameworks for creating linked data in museums, libraries, and archives, and suggested that the similar steps taken by these pioneers in linked data can provide a guide for other institutions.[14]

The IMLS grant-funded LiAM (Linked Archival Metadata) Project at Tufts University published a guidebook and reference model designed specifically for

---

[11] Alemu, et al, "Linked Data for Libraries: Benefits of a Conceptual Shift from Library-Specific Record Structures to RDF-based Data Models," 556.
[12] Jane Greenberg and Eva Mendez, "Linked Open Data for Open Vocabularies and HIVE's Global Framework," *El Profesional de la Informacion,* Vol. 21, No. 3 (2012), 240.
[13] Erik T. Mitchell, "Three Case Studies in Linked Open Data," *Library Technology Report, Vol. 49,* No. 5 (2013), 26-43.
[14] Mitchell, "Three Case Studies in Linked Open Data," 42.

archivists to understand and implement linked data.[15] The project included an environmental scan of current linked data practices "to enable reuse of existing tools and implementations while creating new ones," and to provide a "roadmap for tool and implementation needs" identified by the project.[16] The guidebook provides "an outcome-oriented perspective on the particular benefits that linked data should enable for archivists and archival users," with an emphasis on concrete strategies and guidelines for beginning engagement with linked data from existing archival descriptive practices."[17] The LiAM Project Guidebook intends to inspire the expansion of empirical literature on current linked data practices in archival description, with the goal of lowering barriers to implementing linked data and facilitating its use in archival practice more broadly.

The Archival Approach

Kathy Wisser's statement that "archival is an approach rather than a quality" reflects the understanding that archival description is not a black-and-white practice, but is more often a "rickety bridge" that is crossed carefully through the particular interpretations and deliberate steps made by the archivist.[18] The designation of materials as "archival" is a subjective act and depends on the perspective of the archivist and the

---

[15] Eric Lease Morgan and the LiAM: Linked Archival Metadata Project, *Linked Archival Metadata: A Guidebook* (version 0.99, 23 April 2014), accessed 1 July 2014, http://infomotions.com/sandbox/liam/tmp/guidebook.pdf.

[16] LiAM Project: Linked Archival Metadata, "Prospectus for Linked Archival Metadata: A Guidebook," *LiAM: Linked Archival Metadata Blog*, 1 March 2013, accessed 1 July 2014, http://sites.tufts.edu/liam/deliverables/prospectus-for-linked-archival-metadata-a-guidebook/.

[17] LiAM Project: Linked Archival Metadata, "Prospectus for Linked Archival Metadata: A Guidebook."

[18] Frank Boles, 64th Presidential Address of the Society of American Archivists (presented at the Society of American Archivists Annual Meeting, Austin, Texas, 14 August 2009).

context of their repository. Archivists rely on appraisal and descriptive standards and practices that have developed over many years of professional dialogue, and that provide established, shared guidelines that influence the individual decisions of archivists. The principles of provenance, original order, and *respect des fonds* that guide each archivist are designed to provide a framework for consistency in the field and to best represent the unique nature of archival collections.[19]

The archivist makes decisions about what is of enduring value and then describes materials so that researchers can find and use them. The process of appraisal, arrangement, and description underlines the responsibility of the archivist, who holds a powerful role in deciding what is included, excluded, and made discoverable. Archivists determine the inclusion or exclusion of materials in the historical record, and professional standards guide their actions in the selection, interpretation, description, and representation of collections that shape future access, discovery, and use of materials by researchers. Jennifer Meehan has illustrated that this responsibility is part of the dynamic relationship between the archivist and the materials.[20] Rather than simply observing and illuminating relationships that already exist within and between archival holdings, archivists are actively involved as agents in defining and describing materials as archival, and in creating metadata to represent and provide access to them.

The Evolution of Archival Descriptive Standards

---

[19] Laura Millar, *Archives: Principles and Practices* (New York, NY: Neal-Schuman Publishers, Inc., 2010).

[20] Jennifer Meehan, "Making the Leap from Parts to Whole: Evidence and Inference in Archival Arrangement and Description," *American Archivist* 72 (Spring/Summer 2009), 72-90.

The archival description standards landscape has shifted dramatically in recent decades from an earlier era when description was geared almost exclusively toward bibliographic records describing books, with no appropriate models for archives and manuscripts materials. Standards such as the International Standard for Bibliographic Description (ISBD) and the MARC format provided standardized access points for bibliographic catalog records. However, in 1983 the Library of Congress released the MARC format for Archives and Manuscripts (MARC-AMC), beginning a wave of standards developed with archival description in mind.[21] In the wake of its release, archivists began to recognize that the 1978 edition of the Anglo-American Cataloguing Rules (AACR2), while written in consultation with archivists in the Manuscript Division of the Library of Congress, presented major shortcomings in its application to describing the archival records.[22] The Archives, Personal Papers and Manuscripts (APPM) standard was published in 1983 in response to AACR2, to better address the unique cataloging needs of archival description.

Building upon the foundation set by APPM, the Society of American Archivists developed the current standard in Describing Archives: A Content Standard (DACS) in 1994. In addition to the influence of APPM, DACS takes its core elements from those outlined in the International Standard Archival Description General (ISAD-G). ISAD-G replaced ISBD for archivists by providing an international standard that applies to archival description, and provides a framework for countries to use when developing

---

[21] Steve Hensen, "Squaring the Circle: The Reformation of Archival Description in AACR2," *Library Trends* 36 (Winter 1988), 540.
[22] Hensen, "Squaring the Circle: The Reformation of Archival Description in AACR2," 540-541.

their national standard. DACS is the American standard developed by SAA and provides a framework for describing the content of archival holdings within an institutional context. Building from the international standard of ISAD and the national standard of DACS, archivists create their own internal standard that is documented in a processing manual for their institution, outlining the specific elements of description that are required and how they are to be implemented.

Encoded Archival Description (EAD) was created in 1993, through partnerships between the Library of Congress and the Society of American Archivists, and has evolved as an XML schema for encoding finding aids. EAD represents elements of archival description as outlined in DACS, using a document type definition (DTD), and provides an output standard for producing standardized multi-level archival description in finding aids online. Like the hierarchy of ISAD-G, DACS, and local processing manuals, EAD follows a chain of authority descending from the main EAD standard to a secondary consortium on a more regional or state level, and then is typically customized into a localized template for each repository within that secondary authority's jurisdiction. While in theory EAD is a syntactic data standard, it also operates as a semantic content standard and guides the description of archival holdings following from each DACS element.

Elizabeth Dow argues that the structured data schema of Encoded Archival Description (EAD) does not provide an adequate, sustainable solution for the current needs of archival description.[23] Dow states that EAD is a "halfway technology," akin to

[23] Elizabeth H. Dow, "Encoded Archival Description as a Halfway Technology," *Journal of Archival Organization*, 7:3, (2009), 108-115.

an iron lung, that "[addresses] symptoms of a problem but not the causes or long-term effects."[24] In contrast, she states, a high-level solution akin to the polio vaccine is needed for archivists to treat the deeper problem, which she defines as the need to "connect researchers to archival materials."[25] According to Dow, while EAD achieves its intended purpose by providing archivists with a way to publish finding aids online, it is overly complex, costly, and difficult to use and implement, while providing little long-term reward for the investment of archivists' resources.[26]

In her explanation of Edward Lias's Principle 8 for defining new and old media, which states that, "New Media Causes Underlying Social Values to Change," Dow's observation that "younger researchers seem to work with the assumption that what is not on the Web probably is not worth their time to pursue."[27] Researchers who expect to find exhaustive and immediate results for their search needs may assume that if something is not accessible online then it does not exist. While Dow asserts that EAD and online finding aids are to blame for this user misconception, the speed and comprehensiveness of widely used search functions such as Google have also contributed more broadly to shaping user expectations of the ability to locate and use materials online.

In a 2008 presentation at SAA, Daniel Pitti proposed a solution to concerns surrounding EAD that would combine "markup and database technologies... in complementary ways to form a compete and flexible system of archival description" that would reveal relationships beyond the traditional hierarchical structure, drawing

---

[24] Dow, "Encoded Archival Description as a Halfway Technology," 108.
[25] Dow, "Encoded Archival Description as a Halfway Technology," 110.
[26] Dow, "Encoded Archival Description as a Halfway Technology," 109-110.
[27] Dow, "Encoded Archival Description as a Halfway Technology," 112.

relationships "within a single collection... [and] among various collections that would remain otherwise undetectable."[28] This proposal intends to move archival description toward greater engagement with semantic web technologies and linked data as means to increased interconnectivity, discovery, and use online.

Sheila Morrissey has pointed out issues with consistency of metadata and interoperability within the archival community. She asserts that metadata standards like XML can provide the "benefits of working with a shared vocabulary, applied in accordance with shared practices."[29] This shared vocabulary and standard practice for metadata provide the potentially powerful groundwork for interoperability. However, the "granularity of detail" that is required for encoding content in these metadata standards places a barrier for widespread usability and interoperability. Morrissey points out that the adaptations to the standard that users create as a "solution to an immediate local need," can be a "problem for interoperability," and argues that metadata standards will have to accommodate a culture of "adapters" that will apply the standards according to the terms and practices that best suit their local needs.[30] According to Morrissey, while XML offers a powerful tool for structuring archival metadata, the technical complexity of the rules and standards that govern XML can lead to local adaptations and workarounds, producing more localized silos of practice that create barriers to its broader goal of interoperability.

---

[28] Dow, "Encoded Archival Description as a Halfway Technology," 113.
[29] Sheila M. Morrissey, "More What You'd Call 'Guidelines' than Actual Rules: Variation in the Use of Standards," *Journal of Electronic Publishing*, Vol. 14, Issue 1 (Summer 2011), accessed 1 July 2014, http://dx.doi.org/10.3998/3336451.0014.104.
[30] Morrissey, "More What You'd Call 'Guidelines' than Actual Rules: Variation in the Use of Standards," 8-9.

Heather MacNeil addresses the suggestion, as argued in the growing literature on electronic records management, that metadata systems may render traditional archival description obsolete.[31] According to MacNeil, while metadata systems capture functional and contextual information of records at a granular level to support primary use in their original environment, archival description captures and presents information about records on a broader level to support use of the record by both primary users in the original environment and secondary users in archives. MacNeil argues for the continuing relevance of description, and asserts that traditional methods of archival description remain essential, perhaps more than ever, to describing electronic records in the digital age, in order to preserve electronic records and to make them accessible for researchers in keeping with the established principles of archival practice.[32]

Current Projects Using Linked Data

After observing a growing "appetite for educating the broader community of libraries, archives, and museums about the concepts of Linked Open Data,"[33] in 2011 Jon Voss and others formed the first Linked Open Data in Libraries, Archives, and Museums (LODLAM) Summit, in addition to creating the website LODLAM.net and the community Twitter hashtag #lodlam, as part of efforts to "galvanize international collaboration."[34] The LODLAM Summit continues to meet annually to further Voss's vision of gathering practitioners from different domains of humanities, cultural heritage,

---

[31] Heather MacNeil, "Metadata Strategies and Archival Description: Comparing Apples to Oranges," *Archivaria* 39 (April 1995), 22-32.
[32] MacNeil, "Metadata Strategies and Archival Description: Comparing Apples to Oranges," 27.
[33] Jon Voss, "LODLAM State of Affairs," *Information Standards Quarterly*, Vol. 24, No. 2/3 (Spring/Summer 2012), 41.
[34] Voss, "LODLAM State of Affairs," 42.

and information science, with the goals of "identification of the tools and techniques for publishing and working with Linked Open Data," "drafting of precedents and policy for copyright considerations," and "publishing of definitions and promotion of use cases" that support advocacy for funding linked data projects in libraries, archives, and museums.[35]

Major international libraries have contributed to the development of linked data vocabularies for their resources. The Library of Congress has formed a Linked Data Service that "provides access to commonly found standards and vocabularies promulgated by the Library of Congress," in data sets with resolvable URIs, in order to facilitate programmatic access to authority data at the Library of Congress by humans and machines.[36] The British National Library has also created the British National Bibliography as Linked Open Data platform, which "provides access to the British National Bibliography published as linked data and made available through SPARQL services."[37] The National Archives of Australia has created its AGLS metadata standard for Australian government agencies to create metadata that is compatible with linked data and semantic web applications.[38] The metadata standard operates with both XML and RDF, and is designed to encourage greater discovery and representation of contextual

---

[35] Jon Voss, "Radically Open Cultural Heritage Data on the Web" (presented at Museums and the Web, San Diego, CA, 11-14 April 2012), accessed 1 July 2014, http://www.museumsandtheweb.com/mw2012/papers/radically_open_cultural_heritage_data_on_the_w.

[36] "About," Library of Congress Linked Data Service, Authorities and Vocabularies, accessed 1 July 2014, http://id.loc.gov/about/.

[37] "Welcome to http://bnb.data.bl.uk/," British National Bibliography, accessed 1 July 2014, http://bnb.data.bl.uk/.

[38] "AGLS," National Archives of Australia, accessed 7 July 2014, http://www.naa.gov.au/records-management/agency/create-capture-describe/describe/agls.aspx.

information for users seeking information about government services and agencies in Australia.

The Virtual International Authority File (VIAF), hosted by the Open Computer Library Center (OCLC) and maintained by international libraries including the Library of Congress and national libraries in Europe, maintains a growing number of name authority files online. The service is designed to be "a building block for the Semantic Web to enable switching of the displayed form of names for persons to the preferred language and script of the Web user."[39] In line with the principles of linked data, the creators of VIAF seek to create a "cluster record" or "super" authority record for each entity by linking together national and regional-level authority records. This effort purports to facilitate research across linguistic and local differences in terminology and descriptive conventions, by bringing together all names for an entity into one record.

The Encoded Archival Context Working Group of the Society of American Archivists and the Staatsbibliothek zu Berlin designed Encoded Archival Context: Corporate Bodies, Persons, and Families (EAC-CPF) to accomplish a similar goal in the archival community, by creating a network of authority files for archival creators in order to better preserve the context of archival collection. According to the creators of the 2010 initial release of the standard, "Since materials by or about a single entity might be found in many fonds or many repositories, there is much redundant effort in recording information about the same entity. In addition, these duplicative efforts can result in great

[39] "Virtual International Authority File (VIAF)," OCLC, accessed 1 July 2014, http://www.oclc.org/viaf.en.html.

inconsistency, which bedevils both users, in finding and interpreting materials, and archivists, in creating accurate and complete references to such entities."[40]

The design of EAC-CPF supports the principles of linked data, as its central goal is "to standardize the encoding of descriptions about agents to enable the sharing, discovery and display of this information in an electronic environment."[41] The Social Networks and Archival Context Project (SNAC) is using EAC-CPF records to build a national authority file network of archival entities, with the goal of making historical records more discoverable and useable, in addition to preserving the archival principle of context and respect des fonds by providing "access to the socio-historical contexts (which includes people, families, and corporate bodies) in which the records were created."[42] The SNAC Project is sourcing data for archival authority records from a wide array of established international sources in the United States, the United Kingdom, and France, including the U.S. National Archives and Records Administration (NARA), the Smithsonian Institution, the Library of Congress, the British Library, the Archives Nationales de France, and the Bibliotheque Nationale de France, in addition to linking with descriptions in OCLC Worldcat, VIAF, and the Getty Vocabulary Program.[43]

Europeana is a project funded by the European Commission and hosted in the

---

[40] "EAC-CPF Tag Library," Encoded Archival Context Working Group of the Society of American Archivists and the Staatsbibliothek zu Berlin, version 2010 draft, accessed 1 July 2014, http://www3.iath.virginia.edu/eac/cpf/tagLibrary/cpfTagLibrary.html

[41] "Encoded Context for Archives: Corporate Bodies, Person, and Families (EAC-CPF)," Society of American Archivists, accessed 1 July 2014, http://www2.archivists.org/groups/technical-subcommittee-on-eac-cpf/encoded-archival-context-corporate-bodies-persons-and-families-eac-cpf.

[42] "SNAC Home," SNAC: Social Network and Archival Context Project, accessed 1 July 2014, http://socialarchive.iath.virginia.edu/.

[43] SNAC Project, "SNAC Home."

National Library of the Netherlands that harvests metadata about digital objects from cultural heritage institutions across Europe, and has created a single search portal that facilitates search and discovery of materials across archival, museum, and library collections.[44] Europeana defines Linked Open Data as "a way of publishing structured data that allows metadata to be connected and enriched, so that different representations of the same content can be found, and links made between related resources."[45] Research participant Valentine Charles, Data and Research and Development Coordinator at Europeana, explained that Europeana relies on linked data as part of its Europeana Data Model in order to pursue its goals of creating a "distribution channel" to cultural heritage resources online, of preserving the context of those resources, and of facilitating reuse of the data that represents those resources.[46] The Europeana Data Model utilizes the existing framework of linked data to "describe the rich resources from museums, libraries, archives, and audiovisual archives, and then link this data to existing datasets such as DBpedia (a data version of Wikipedia) or Geonames (structured geographical data)."[47] Europeana and its data model have guided other projects such as the Digital Public Library of America in similar efforts to harvest metadata and create unified search portals for library, museum, and archival resources online.[48]

While the Europeana Data Model has focused primarily on description of digital

---

[44] "Portal Homepage," Europeana, accessed 1 July 2014. http://europeana.eu/portal/.
[45] "Thought Lab," Europeana, accessed 1 July 2014, http://pro.europeana.eu/web/guest/linked-open-data.
[46] Valentine Charles, interview with Virginia Ferris, 10 April 2014.
[47] Europeana, "Thought Lab;" "Europeana Data Model (EDM) Documentation," Europeana, accessed 1 July 2014, http://pro.europeana.eu/edm-documentation.
[48] Valentine Charles, interview with Virginia Ferris, 10 April 2014; "Welcome," Digital Public Library of America, accessed 1 July 2014, http://dp.la/.

objects, Europeana appointed a Task Force on Hierarchical Objects to publish recommendations for representing hierarchical resources such as archival collections using Europeana's EDM descriptive framework and user interface.[49] The Task Force's report relied on case studies for analysis of issues for hierarchical description, including studies of family papers, artist's papers, and other archival collections, in addition hierarchical objects in archaeological, audiovisual, and other types of collections.[50] The report outlines best practices and recommendations for confronting various challenges. Highlighted challenges include the importance of preserving contextual relationships, inconsistent availability of contextual information, the need for inheritance of metadata between different levels of hierarchy, search and navigation between different levels, different levels of granularity of hierarchy description, the prevalence of collection materials lacking a digital representation in archival collections that are not digitized, and the greater question of how to display hierarchical objects.[51] Recommendations center upon updating the Europeana Data Model to make "small adjustments" and advise greater "guidance in terms of content strategy" for hierarchical objects.[52]

The Digital Repository of Ireland has piloted several collaborative projects, including the Linked Logainm: Location LODer project, which provides "an authoritative

---

[49] "Task Force on Hierarchical Objects," Europeana, accessed 1 July 2014, http://pro.europeana.eu/web/network/europeana-tech/-/wiki/Main/Taskforce+on+hierarchical+objects.
[50] "Recommendations for the Representation of Hierarchical Objects in Europeana," Europeana Task Force on Hierarchical Objects: 5-6, accessed 1 July 2014, http://pro.europeana.eu/documents/468623/4a6eb2ec-4cc6-48b1-8824-92a1e564a279.
[51] Europeana Task Force on Hierarchical Objects, "Recommendations for the representation of hierarchical objects in Europeana," 39-40.
[52] Europeana Task Force on Hierarchical Objects, "Recommendations for the representation of hierarchical objects in Europeana," 21-27.

database of Irish placenames" in English and Irish Gaelic, presented with a Google maps interface and using linked data from the National Library of Ireland, DBpedia, Europeana, and the Irish Historic Town Atlas.[53] The Linked Logainm project intends to be "widely used by web developers, computer scientists, researchers, the heritage community, and librarians and archivists."[54] Other projects include Irish Record Linkage, which intends to "provide a comprehensive map of infant and maternal mortality for Dublin from 1864-1913," by reusing digitized "legacy data" from government records "to reconstitute families and create longitudinal histories by linking birth, death and marriage (BDM) records" to create a "knowledge platform" that uses the principles of linked data to allow researchers to "interrogate this historical Big Data."[55]

Archives Hub provides another example of a leading European project in linked data, in this case focusing specifically on aggregating descriptions of archival collections across the United Kingdom.[56] Archives Hub is a Joint Information Systems Committee (JISC) funded service based in Mimas at the University of Manchester. The service emphasizes collaboration with archivists to provide a "gateway to thousands of the U.K.'s richest archives" for the use of scholars, educators, and broader research use among the general public.[57] Jane Stevenson, Manager of Archives Hub, and Adrian Stevenson, of Mimas at the University of Manchester, collaborated in the development of

---

[53] "Projects," Digital Repository of Ireland, accessed 1 July 2014, http://dri.ie/projects; "Linked Logainm: Placenames Database," Digital Repository of Ireland, accessed 1 July 2014. http://www.logainm.ie/en/; "LocationLODer," Digital Repository of Ireland, accessed 1 July 2014, http://apps.dri.ie/locationLODer/.
[54] Digital Repository of Ireland, "Location LODer."
[55] "Projects," Digital Repository of Ireland, accessed 1 July 2014, http://dri.ie/projects.
[56] "Home," Archives Hub, accessed 1 July 2014, http://archiveshub.ac.uk/.
[57] "Introducing the Archives Hub," Archives Hub, accessed 1 July 2014, http://archiveshub.ac.uk/introduction/.

"Linking Lives," a pilot end-user interface based on the Linked Open Copac and Archives Hub (LOCAH) project.[58] Jane Stevenson contributed an evaluation report of Linking Lives as a test case for determining user needs and reactions to a linked data interface for archival discovery. Stevenson's report found that users were concerned about the provenance and quality of data, doubting the trustworthiness of data sets from sources such as Wikipedia; additionally, users were concerned with how and why data was selected, and with the comprehensiveness and authoritativeness of selected data sources.[59] Stevenson acknowledged:

> "The benefit of Linked Data is that you can draw in a diverse set of sources, and the aim is to provide a well-rounded view, but the more sources you pull into a single interface, the more you have to consider how to present them clearly; to show that they are distinct sources and to convey to the researcher that they are not under your control."[60]

Stevenson concludes that an important issue for archival linked data projects moving forward will be to investigate the design of interfaces and search mechanisms to present linked data to users in a logical, meaningful "way that really answers their research questions."[61] Stevenson advocates for continuing user testing in the development of linked data services.[62]

In 2013 the Australian Research Council funded the Linked Data Policy Hub project, which seeks to "transform the accessibility and utility of social science

---

[58] "About," LOCAH Linked Archives Hub, accessed 1 July 2014, http://data.archiveshub.ac.uk/about.html.
[59] "Linking Lives Evaluation Report," Archives Hub, accessed 1 July 2014, http://archiveshub.ac.uk/linkinglives/.
[60] Archives Hub, "Linking Lives Evaluation Report."
[61] Archives Hub, "Linking Lives Evaluation Report."
[62] Jane Stevenson, "'What Happens If I Click on This?': Experiences of the Archives Hub," *Ariadne,* no. 57 (October 2008), accessed 1 July 2014, http://www.ariadne.ac.uk/issue57/stevenson.

publications, grey literature and data in Australia by linking documents and data from multiple collections across the internet, enabling innovative new forms of social and policy research and analysis."[63] The project will created linked data sets for two major research collections, Australian Policy Online and Informit, and connect them using controlled vocabularies to create "online hubs of linked resources designed for research across the social sciences."[64] While the project focuses on social science research data, its goals align with other international linked data projects for archival and cultural heritage by seeking to enable researchers to "analyse and visualise relationships between documents and data sets from a wide range of sources in order to better understand underlying patterns and trends."[65]

The "Invisible Australians" project, developed by Tim Sherratt and Kate Bagnall, seeks to utilize facial recognition technology and linked data to connect photographs with other government documents kept in the National Archives of Australia, in order to interrogate historical narratives surrounding race and national identity in Australia.[66] Sherratt explains that the digitized collections of records in the National Archives of Australia facilitated his development of experimental interface that displays images with links to government documents, signaling a shift in power between repositories, developers, and users: "By putting part of the collection online, they provided us with the

---

[63] "Linked Data Project," Australian Policy Online, accessed 7 July 2014, http://apo.org.au/content/linked-data-project.
[64] Australian Policy Online, "Linked Data Project.
[65] "Linked Data Policy Hub – connected resources for social research," Research Data Australia, accessed 6 July 2014, http://researchdata.ands.org.au/linked-data-policy-hub-connected-resources-for-social-research.
[66] "Home," Invisible Australians: Living Under the White Australia Policy, accessed 6 July 2014, http://invisibleaustralians.org/.

opportunity to develop a resource that both extends and critiques the existing collection database. Interfaces to cultural heritage collections are no longer controlled solely by cultural heritage institutions."[67] Sherratt presents the perspective of developers who are using open source tools to develop alternative representations of cultural heritage materials, facilitating new avenues for discovery and use archival records online for the broader public that purposefully creates opportunities for "finding the oppressed, the vulnerable, the displaced, the marginalized and the poor and giving them their place in history."[68]

In response to growing interest and projects using linked data in the library, archives, and cultural heritage community, there are a growing number of tools and best practices emerging to assist interested institutions in publishing linked data. OpenRefine, an extension of the former GoogleRefine, provides a set of tools as part of the Free Your Metadata initiative to help guide institutions through the process of making their data ready for sharing as linked data. The initiative is product of collaboration between the Multimeda Lab at Ghent University and MasTIC at the Université Libre de Bruxelles. Free Your Metadata focuses primarily on museums, but instructs any interested institutions in their prescribed process using OpenRefine: "cleanup," providing a checklist of how to find and correct errors in data in a semi-automated way;[69] "reconciliation," showing how to connect collection-specific vocabularies to more widely

---

[67] "It's all about the stuff: Collections, interfaces, power, and people," Tim Sherratt, Invisible Australians: Living Under the White Australian Policy, accessed 6 July 2014, http://invisibleaustralians.org/blog/2011/12/it%E2%80%99s-all-about-the-stuff-collections-interfaces-power-and-people/.

[68] Tim Sherratt, "It's all about the stuff: Collections, interfaces, power, and people."

[69] "Cleanup," Free Your Metadata, accessed 1 July 2014, http://freeyourmetadata.org/cleanup.

shared controlled vocabularies and "making your metadata interpretable by the whole wide world;"[70] "extraction," in capturing named entities from unstructured free text and extracting it in a machine-processable format;[71] and "access," which, once developed, is expected to provide a series of guidelines for publishing "liberated" metadata in a sustainable way, "illustrated by a real-world use case."[72]

The W3C Schema.org for Bibliographic Extension (SchemaBibEx) Community Group was founded by Richard Wallis of OCLC in 2012 as a community of technicians exploring the potential of "seeding" Schema.org into existing library systems to create structured data, as a solution to the "barrier imposed by implementation constraints" for linked data.[73] The mission of the SchemaBibEx community is "to discuss and prepare proposal(s) for extending Schema.org schemas for the improved representation of bibliographic information markup and sharing."[74] According to Scott, using Schema.org vocabularies along with RDFa, microdata, and JSON-LD formats, would allow library catalogues to "feed metadata directly to general search engines," taking library data out of silos of library-specific systems that prevent widespread discovery on the web, and

---

[70] "Reconcile," Free Your Metadata, accessed 1 July 2014, http://freeyourmetadata.org/reconciliation.
[71] "Extraction," Free Your Metadata, accessed 1 July 2014, http://freeyourmetadata.org/named-entity-extraction.
[72] "Access," Free Your Metadata, accessed 1 July 2014. http://freeyourmetadata.org/sustainable-access.
[73] Dan Scott, "Structured Data NOW: Seeding schema.org in library systems," abstract in *Code4Lib* 2014 conference program, accessed 1 July 2014. http://code4lib.org/conference/2014/scott.
[74] "About," W3C Schema Bib Extend Community Group, accessed 1 July 2014, http://www.w3.org/community/schemabibex/.

enabling libraries to engage with linked data with little disruption to existing systems.[75]

The Decipher project is a partnership of several Irish cultural heritage institutions, the National Gallery of Ireland (NGI) and the Museum of Modern Art Ireland (IMMA), with developers and academic research centers in Ireland, funded by the European Union 7th Framework Programme in the area of Digital Libraries and Digital Preservation.[76] Decipher is committed to utilizing semantic web technologies for cultural heritage resources, particularly in museum settings. The project emphasizes the central importance of narrative structures in representing relationships, in order to communicate and understand the meaning of cultural heritage materials: "Cultural meaning does not reside in individual objects but in the patterns of knowledge and events, belief and thought that link them to each other and to the observer."[77] Building from this perspective, the Decipher project seeks to create:

> "A reasoning engine, virtual environment and interfaces that can present digital heritage objects as part of a coherent narrative, directly related to individual searches and user contexts. This will allow the user to interactively assemble, visualise and explore, not just collections of objects, but the knowledge structures that connect and give them meaning."[78]

Decipher is currently supporting the development of Curate, an ontology that is currently being used with Storyspace to model museum narratives.[79]

---

[75] Dan Scott, "Structured Data NOW: Seeding schema.org in library systems" (slides presented at *Code4Lib*, Raleigh, NC, 2014), accessed 1 July 2014, http://stuff.coffeecode.net/2014/structured_data_now/#/8.
[76] "Introduction," Decipher Project, accessed 1 July 2014, http://decipher-research.eu/introduction.
[77] Decipher Project, "Introduction."
[78] Decipher Project, "Introduction."
[79] Paul Mulholland, "Curate: An Ontology for Describing Museum Narratives," *Decipher Project*, 1 November 2013. Accessed 1 July 2014. http://decipher.open.ac.uk/curate/introduction; Annika Wolff, Paul Mulholland, and

Eoin Kilfeather of the Decipher project echoes Jane Stevenson's conclusions, in writing that there is a "tension between providing user control over what is shown and the coherence of the overall experience" of cultural heritage materials displayed online using linked data.[80] While Kilfeather and the Decipher project are concerned primarily with museum collections, archival collections face many of the same issues with preserving dynamic contextual relationships:

> "Cultural narratives are not static, but change in time and in relation to the viewpoints from which they are both created and explored. Curators have developed ways of presenting their collections that guide visitors and make sense of the experience. Although we have new ways of accessing cultural objects, the narrative structures and arguments that could be found in a handcrafted presentation are often lost."[81]

The literature suggests that creating URIs and establishing relationships for cultural heritage data using linked data could provide a way to structure information in a way that would allow the evolution of historical and contextual relationships to occur, while preserving the fundamental provenance and networks of relationships that are critical to retaining meaning in archival collections.

Trevor Collins, "Storyspace: A Story-Driven Approach for Creating Museum Narratives" (Milton Keynes, United Kingdom: Knowledge Media Institute, the Open University, 2012), accessed 1 July 2014, http://people.kmi.open.ac.uk/paulm/hypertext-2012.pdf.

[80] Eoin Kilfeather, "Deciphering the Story of the Museum," *ERCIM News*, July 2011, accessed 1 July 2014, http://ercim-news.ercim.eu/en86/special/deciphering-the-story-of-the-museum.

[81] Kilfeather, "Deciphering the Story of the Museum."

METHODOLOGY

The researcher collected data through structured interviews with practitioners familiar with linked data and archival description, and through an environmental scan of institutions and projects currently working with linked data and archival description. Data collection methods were influenced by Jesse F. Brown's qualitative study of early adopters of the Encoded Archival Context (EAC) authority control standard.[82] Following an environmental scan of institutions and projects involved with the use of linked data, metadata, archival description, and descriptive standards, the research identified key projects including Europeana, Archives Hub, the Tufts University LiAM Project, and Social Networks and Archival Context (SNAC) Project, and contacted individuals associated with these projects as staff or advisory board members. Interview subjects were invited to participate through email. An initial email was sent to determine interest, and for each of the interested respondents the researcher sent a copy of a consent form through email and asked respondents to return a signed and completed consent form to the researcher via email. In the consent form each subject indicated whether they wished to be identified, along with their institution, in the study. Following the researcher's receipt of consent forms, a second email was sent to schedule a recorded telephone or Skype video chat interview.

---

[82] Jesse Brown, "More Than Just a Name: Archival Authority Control, Creator Description, and the Development of Encoded Archival Context (EAC)" (master's paper, School of Information and Library Science, University of North Carolina at Chapel Hill, 2006).

While participants were given the option to remain anonymous, all but one participant chose to be identified. The following participants chose to be identified in the study by name, job title, and institution:

- Eric Lease Morgan, Digital Projects Librarian, Notre Dame University.
- Valentine Charles, Data and R&D Coordinator, Europeana.
- Corey Harper, Metadata Services Librarian, New York University.
- Kathy Wisser, Assistant Professor, Simmons College.
- Daniel Pitti, Associate Director, Institute for Advanced Technology in the Humanities, University of Virginia.
- Jane Stevenson, Manager, Archives Hub.
- Adrian Stevenson, Library and Archival Services, Technical Innovations Coordinator, Mimas, University of Manchester.

Subjects participated in 30-minute interviews that were conducted over the telephone or Skype video chat in a semi-structured style. Each interview was recorded on a digital voice recorder and transcribed for data analysis. The researcher asked a standardized list of questions for all interview subjects.[83] Questions were adjusted individually to focus on the use of linked data in the context of each subject's institution, and based on the subject's level of familiarity or length of time that the subject has used linked data practices; individuals with more experience with archival description and archival authority work were asked more targeted questions in these subject areas, while participants with more experience in linked data and semantic web practices were asked more detailed questions in these areas. Interview questions were developed to assess

[83] See Appendix A for list of interview questions.

professional attitudes toward linked data and perceived impact of its use on archival description, locally and globally.

Following the recording and transcription of interviews, each interview was coded and outlined according to topics related to specific research questions using a list of subject headings. The researcher developed subject headings to identify major statements made by each participant in relation to the study's central research questions.[84] After applying subject headings, the researcher conducted analysis of the responses, to provide thick description of professional attitudes, experiences, and applications of linked data for online description of archival materials. Together with the environmental scan of current projects using linked data, the interviews will provide insight into the experiences of archivists and metadata specialists working with linked data, to bring into focus the current state of linked data perceptions, applications, and recommendations within the archival community.

[84] See Appendix B for list of interview subject heading codes.

FINDINGS

Interviews with research participants focused on several major topics. First, the researcher gathered participants' definitions of linked data and information about their background in learning about and using linked data. Second, participants discussed their perceptions of the potential value of linked data for users and practitioners in archives. Third, participants discussed their perceptions of the current barriers and challenges in implementing linked data and semantic web approaches in archival description. Fourth, participants offered their perceived solutions for potentially removing the current barriers to applying linked data in an archival context. Finally, participants discussed their greater conclusions and predictions regarding future patterns and developments in the areas of archival description and linked data.

Defining Linked Data

The first question posed to participants was: How do you define linked data? Responses fell along a range of technical specificity, with some respondents defining linked data according to Tim Berners-Lee's definition, realized through RDF triples, and others defining linked data as more of a conceptual framework, and most fitting somewhere in between these two statements.

Eric Lease Morgan of Notre Dame University stated that linked data is "a way to publish information about things… [or] entities" on the web in a way that every entity has some description or metadata associated with it using a URI. Morgan acknowledged that, "there are different ways that you can exposed linked data," but the general rule is to

expose this data through RDF triples to promote interoperability and to "allow other systems that can work with RDF to link to those things."[85]

Corey Harper of New York University defined linked data first as a conceptual model that is "moving away from a record-based model for managing metadata and towards thinking about metadata resource description as a graph," and explained that the way to realize and enact this model is through the principles defined by Tim Berners-Lee:

> "Using URIs for the names of things, using HTTP URIs so that people can look things up, giving back information when people look those things up, and providing the links to other URIs so that people can follow them around and find other interesting stuff on the web."[86]

Kathy Wisser of Simmons College defined linked data as a conceptual framework for description that centers on "leveraging the description of relationships between different data points," and noted that archivists have "been [leveraging the description of relationships] for a really long time, it's just that we haven't called it linked data before." According to Wisser:

> "Archival description is all about linked data. If you think about it not in the context of explicitly accepting it in RDF triples, but in the context of establishing relationships between information, or between things, that's what archival description does. It says that this group of stuff has a relationship with each other because it's in the same format or it covers the same topic or it was created at the same time period or by the same author. So we've been doing linked data since we've been doing archival description."[87]

Jane Stevenson of Archives Hub defined what she termed "five-star linked data" according to the technical framework outlined by Berners-Lee. Five-star linked data typically refers to data that is: 1) On the web, 2) Machine-readable, 3) Non-proprietary

---

[85] Eric Lease Morgan, interview with Virginia Ferris, 28 March 2014.
[86] Corey Harper, interview with Virginia Ferris, 23 April 2014.
[87] Kathy Wisser, interview with Virginia Ferris, 8 May 2014.

format, 4) RDF Standards, 5) Linked RDF. Stevenson distinguished this five-star linked data from more abstract or conceptual notions of linked data, and noted that there is some confusion about what linked data means: "People think if in some way they link data they've created linked data."[88] Stevenson noted that in her work with Archives Hub she maintains a strong emphasis on including external links to other data sets, in keeping with the fullest expression of linked data:

> "I sometimes think people create RDF versions of their data, but they don't prioritize the links into other data sets, and if you don't do that – as far as I'm concerned – we're not going to realize the vision of linked data."[89]

Adrian Stevenson, also of Archives Hub, referred to the range of understandings of linked data along from the technical to the conceptual as the different between "upper case Linked Data or lower case linked data." Stevenson echoed the group's commitment to Berners-Lee's principles of five-star linked data, noting that "there are looser notions, but we tend to honor that general standard."[90]

Daniel Pitti of the University of Virginia and of the International Council of Archivists referred to the different definitions of linked data, first in the technical understanding that "you have a machine-actionable link between two resources that are on the Internet, and that you use URIs or URLs to uniquely identify their resources, and also uniquely identify the relation that exists between them."[91] He stated that this framework "opens up a world of possibilities in terms of making statements about resources and chaining these together in a series of statements that are interlocked with

---

[88] Jane Stevenson, interview with Virginia Ferris, 30 April 2014.
[89] Jane Stevenson, interview with Virginia Ferris.
[90] Adrian Stevenson, interview with Virginia Ferris, 30 April 2014.
[91] Daniel Pitti, interview with Virginia Ferris, 6 May 2014.

one another." Pitti's observation of the potential for linked data to represent relationships between resources in a structured way bridges the technical and conceptual frameworks of linked data.

Exposures to Linked Data

After determining participants' perspectives on defining linked data, participants were asked to describe their own experiences learning and working with linked data. Most participants described first learning about linked data around the time it was introduced by Tim Berners-Lee in the late 1990s and early 2000s. Corey Harper became interested in linked data early in his career while a graduate student at the University of North Carolina at Chapel Hill, through his work with Jane Greenberg and her metadata research center. Harper connects the early development of linked data with parallel developments in Dublin Core metadata standards:

> "The Dublin Core community actually has its roots in a lot of the same places as the linked data community, so when the semantic web was first taking off in the mid 90s, a lot of the same people who were doing Dublin Core were also doing specifications around the Resource Description Framework. So I got inundated or indoctrinated with this stuff pretty early, basically in the onset of my career."[92]

One respondent stated that he has learned most from informal sources such as blogs and listservs, as well as articles published online and in the Code4Lib journal."[93] Others acknowledged what they felt were the "hyperbole" or "overselling" of linked data and the semantic web in its introduction, noting that while linked data has become a popular concept, there are many issues and challenges that have not yet been resolved.[94]

---

[92] Corey Harper, interview with Virginia Ferris, 23 April 2014.
[93] Anonymous research participant #992-02, interview with Virginia Ferris, 3 April 2014.
[94] Anonymous research participant #992_02, interview with Virginia Ferris; Daniel Pitti, interview with Virginia Ferris, 6 May 2014.

Several participants remarked upon the challenge of making pragmatic steps in applying linked data after establishing a strong knowledge base. Valentine Charles of Europeana explained that early in the development of the Europeana Data Model and extensive work with linked data, while the staff had "quite a good understanding of what [linked data] meant," their greatest challenge was "putting it into practice."[95] Charles noted:

> "I won't say we knew everything about linked data, but we were quite in a good position… and we had practical knowledge of what is feasible, what technology is available, and the vocabularies. But then there's matching it to the reality of what you have… It's not a question that we didn't know enough, it's just applying it to the reality of what you have."[96]

Many agreed that the key to breaking through the challenges of putting knowledge about linked data into practice lies in experimentation. Corey Harper stated that in his work with linked data, "It's all been experimental."[97] Harper took a research sabbatical from his work at New York University to create a pilot system for transforming EAD into linked data, and states that while his developments were not directly implemented at NYU, the principles of linked data are influencing strategic initiatives and "forming our systems architecture and our mid- to long-term metadata management strategies."[98]

Kathy Wisser explained that while she has not worked closely with linked data in the technical RDF sense, she consults on projects related to the concept of linked data and its potential applications in the form of the Tufts LiAM (Linked Archival Metadata), for

---

[95] Valentine Charles, interview with Virginia Ferris, 10 April 2014.
[96] Charles, interview with Virginia Ferris.
[97] Corey Harper, interview with Virginia Ferris, 23 April 2014.
[98] Harper, interview with Virginia Ferris.

which Eric Lease Morgan published the recent *Linked Archival Metadata Guidebook.*[99] Wisser positioned her concerns regarding linked data as rooted in its conceptual framework and potential impact on archival description, and outlined her advocacy for "meaningful linked data":

> "I'm not coming to it from the technology end of it, I'm coming at it from the philosophical end and archival description end of it… I'm actually a big fan of meaningful linked data… I think that we should link things that should be linked, and I think that we need to think very carefully about that. What does that actually mean? If we're actually going to invest some meaning into it and add value, let's really explore what it means to link things and which things really need to be linked in order to be useful."[100]

Potential value

After outlining preliminary observations based on each participant's personal experiences with linked data, participants were asked to describe what they viewed as the potential value of linked data for archives. Many respondents noted that linked data could "de-silo" separate domains and institutions, reducing reliance on local standards of archival repositories or distinct standards and practices in different cultural heritage domains. Respondents noted that this "de-siloing" effect would facilitate greater discovery and use of archival materials across archival collections, repositories, and in the broader domains of libraries, museums, and cultural heritage institutions. This would encourage greater visibility and prominence of archives on the open web with other cultural heritage communities. Several respondents commented:

> "In terms of ways that are unique to archives, I think part of the benefit will be elevating the archival resources to becoming first-class citizens in this larger

---

[99] Eric Lease Morgan and the LiAM: Linked Archival Metadata Project, *Linked Archival Metadata: A Guidebook* (version 0.99, 23 April 2014), accessed 1 July 2014, http://infomotions.com/sandbox/liam/tmp/guidebook.pdf.

[100] Kathy Wisser, interview with Virginia Ferris, 8 May 2014.

> collection of cultural heritage materials… [by allowing them to be] integrated into primary discovery streams."[101]
>
> "[Linked data allows one to use] generalized technical formats so the information is not trapped in a niche area using archival standards or, I don't know, geographical data standards or other domains. It's bringing it into a more generalized space where you can hook data together more straightforwardly."[102]
>
> "We become a part of all information sources and all data sources, and it's not anymore about archives or cultural heritage, it's everything. There are no boundaries."[103]
>
> "It tears down the boundaries between the isolated systems in which the data exists and allows it to be taken outside of it and allows things to be done with it. So closed systems can suddenly be exposed and made open so that people can use that data… it has a great deal of potential in terms of breaking down boundaries between cultural heritage communities, which is to say libraries, archives, and museums, to the extent that there is overlap, and there is fairly substantial overlap in the way professionals describe cultural objects, that there is substantial potential to interrelate those systems to provide unified views into the data, or integrated pathways into the distributed and variously described resources."[104]

Several other participants also noted that the de-siloing of cultural heritage institutions through linked data would allow archives to better demonstrate the value, relevance, and richness of their collections:

> "The archival data are really rich in terms of entities – place names and people's names… [With linked data] they could show, 'Yeah, archives' data are really rich and you can do a lot of stuff with it.' By joining the linked data movement [archivists] could start linking their data sets with other data sets and start to show the value of archival materials when they are put close to something else, such as a painting or an object in a museum or a book in a library. Because I guess a lot of people say 'Archives are just paper. It's not interesting.' But maybe what linked data allows you to do is to put things together and start to put things into more context, and archives will be the ones bringing this context."[105]

---

[101] Corey Harper, interview with Virginia Ferris, 23 April 2014.
[102] Adrian Stevenson, interview with Virginia Ferris, 30 April 2014.
[103] Jane Stevenson, interview with Virginia Ferris, 30 April 2014.
[104] Daniel Pitti, interview with Virginia Ferris, 6 May 2014.
[105] Valentine Charles, interview with Virginia Ferris, 10 April 2014.

Kathy Wisser commented upon linked data as a tool that could more fully grant archivists the authority to represent relationships and contextual information, in keeping with core principles of archival practice. She outlines the importance of contextual information in archival description:

> "It gives us [archivists] more power to contextualize the record. And the more we can contextualize the record, the better an understanding of those records will be for researchers… I do see the potential to leverage better contextual information. So if for instance you were looking at a subordinate unit in a corporate body, their records, being able to very quickly get to a description of the larger corporate body and the records that are created by that larger corporate body, and understanding how this unit worked within that larger unit, gives you a better sense of the records that they're actually creating. I work for a subordinate unit of a corporate body. I work for a school within a college, and obviously a lot of decisions that are made at my school have a significant amount to do with the college itself, and the philosophy of the college, and initiatives the college has. If you don't have the ability to make that easy transition, then understanding the school is diminished and the records of the school make less sense."[106]

Others remarked that linked data would "add value to archives" by describing archival collections in a way that could more effectively communicate their potential use and relationships with other materials: [107]

> "An archivist will be able to tell a much more thorough story about their collection if they use linked data principles. They will be able to tell a much richer story**.** It will go beyond, 'This is what I've got.' Yes, it's great that they write those biographical histories and the scope notes about the collection. That's wonderful. But those things will be able to be improved and enhanced and made much more multimedia-rich if they use linked data. That's where the potential comes from."[108]

Several respondents also commented on the potential for linked data to inspire a movement toward improving the quality and consistency of data in archival description. Corey Harper observed that, "Archival practice often creates data without realizing or

---

[106] Kathy Wisser, interview with Virginia Ferris, 8 May 2014.
[107] Eric Lease Morgan, interview with Virginia Ferris, 28 March 2014.
[108] Morgan, interview with Virginia Ferris.

being cognizant of the fact that they are creating data… I don't think people are really thinking that what they're doing is creating metadata. I think people think that what they're doing is document markup, which they also are."[109] Similar to Harper, those who have worked with linked data and archival metadata noted that the increased prominence of linked data has influenced a shift toward thinking more seriously about creating archival description as consistent, well-formed metadata, in order to effectively engage with linked data and its potential benefits:

> "Data is finally becoming more public, so people are taking a bit more care of it."[110]

> "I think one of the benefits for archivists is that we've had to go back and look at our data and think about how consistent it is. And actually in a lot of ways we've had to improve the source data. When you're creating linked data, because you're structuring it so much more, you have to be more rigorous in certain ways… [Linked data] has helped hugely with our thinking about the structure of our data, the consistency and quality of our data, which has been really beneficial."[111]

One respondent observed that linked data presents the potential to save time and create more efficient description workflows for archivists, by sharing and building upon existing description in the form of URIs:

> "I think from the archivist's perspective, it has the potential to save a ton of time. When we do metadata work, I'd say the majority of the values that we're sticking in these records, we're looking up some string in some other database and we're basically copying that string into our own local record, and we're basically just replicating work over and over and over again. The idea with linked data is you don't need to copy that thing. You just point to it. So if we're able to make all that happen, then we have the potential to save a ton of work for catalogers. I mean you still have to look up things, but you're only maintaining the data in one

[109] Corey Harper, interview with Virginia Ferris, 23 April 2014.
[110] Adrian Stevenson, interview with Virginia Ferris, 30 April 2014.
[111] Jane Stevenson, interview with Virginia Ferris, 30 April 2014.

> place. You're not maintaining a hundred different copies of the same subject heading that may exist in thousands of databases."[112]

Several participants commented on the potential for linked data to inspire and facilitate creative reuse of data in creative ways, beyond what archivists can currently predict:

> "I might use the data to create a finding aid, but somebody else might use it to build a report of some sort of computer program that makes a story of some sort, that you didn't imagine."[113]

Daniel Pitti remarked in particular on the openness and versatility of linked data as a key advantage:

> **"**Versatility in terms of making expressions. It most certainly is the ability to take data that's developed in diverse environments and expose it so that third parties can come along and take that data and interconnect it and interrelate with data, and take it from other sources and do things with it that's completely unanticipated by the people who created the data to begin with."[114]

Challenges and barriers

Following this description of the perceived benefits of linked data, participants were asked to describe what they viewed as the greatest potential challenges and barriers to implementing linked data for archives. The most common response among interviewees centered on both the perceived and realized technical complexity of linked data as an obstacle in its adoption. Respondents observed that the concept of linked data

---

[112] Anonymous research participant #992-02, interview with Virginia Ferris, 3 April 2014.
[113] Eric Lease Morgan, interview with Virginia Ferris, 28 March 2014.
[114] Daniel Pitti, interview with Virginia Ferris, 6 May 2014.

is often viewed as abstract, complicated with a "steep learning curve,"[115] and "for the birds,"[116] which decreases its likelihood of adoption and use among archivists:

> "[Linked data] is simply more computer stuff and they'll get tired and frustrated with all that. 'Yet another thing that I have to learn.' Which I completely understand."[117]
>
> "It's just technically complicated, and it's probably beyond the comfort level of most archivists. If you're not working with a tool that makes it really easy to work with linked data, you don't have to necessarily be proficient in writing XML or XSL, or dealing with XML or JSON or some other kind of data structure."[118]
>
> "I think the technology itself is still somewhat young. It's rather complicated and it's not widely understood and appreciated, so there's a social dimension to it in a sense of helping the community to understand and appreciate what it can do with it."[119]
>
> "Although it's a generalized format, I suppose the bit where it comes back to being a bit niche again, is the whole thing of what vocabularies and ontologies to use. There are loads and loads of vocabularies specific to linked data and XML and things like that. In the library community, knowing where to find them, which ones to use, using them correctly, there's quite a lot of effort there. It's not easy to see how that will become mechanized or machinized, if you like. It still seems like a job where you have to apply a lot of human effort… It's quite a lot of information that you probably need to absorb before you can get your head around that."[120]

Building on discussions of the technical complexity of linked data and the traditional descriptive practices of archivists, respondents noted that a major barrier in the adoption of linked data is the relative lack of tools for practitioners. Many agreed that there is a great need for more user-friendly tools that will allow archivists to use and understand linked data, and that can demonstrate the potential uses and benefits of linked

---

[115] Jane Stevenson, interview with Virginia Ferris, 30 April 2014.
[116] Eric Lease Morgan, interview with Virginia Ferris, 28 March 2014.
[117] Morgan, interview with Virginia Ferris.
[118] Anonymous research participant #992-02, interview with Virginia Ferris, 3 April 2014.
[119] Daniel Pitti, interview with Virginia Ferris, 6 May 2014.
[120] Adrian Stevenson, interview with Virginia Ferris, 30 April 2014.

data within the archival community. Several participants commented that while some archivists are interested in using linked data, they do not know how to begin using it without a tool or guide to doing it:

> "I think the challenge of linked data for the archival community, number one, is simply as a lone archivist or an archivist in a small library, it's incredibly difficult to know where to start and to know how to utilize this – how you create linked data, what you do, how you go about it. One of the most common questions that we get is, 'Well where do I start?' You can read a lot about it, but somehow you still sit there at your desk and go, "Okay, what do I do now?" So I think that's really difficult."[121]

> "There are no cool tools that I have on my desktop that I can sit there and fill in a form that says 'Joe Schmoe,' and it comes back and says "This is the URI for Joe Schmoe.' We need tools like that."[122]

> "Most of the tools that archivists use in the course of their daily work aren't built in such a way that is easy to publish linked data or access it. That's changing a little bit."[123]

The second most common barrier cited by respondents was the linear, hierarchical structure and document focus of traditional archival description, as represented in the finding aid. Some observed that the finding aid is traditionally understood and composed as a document, while linked data composed of RDF triples would consist of piles of many statements:

> "Archival description by its nature is just very complicated, compared to description in other library domains. The multi-level description is really unique to archives, and I think it has been really hard to build user-friendly and standard interfaces to archival description because of the way it's marked up. Because it's so complicated, it's been inconsistently done."[124]

---

[121] Jane Stevenson, interview with Virginia Ferris, 30 April 2014.
[122] Eric Lease Morgan, interview with Virginia Ferris, 28 March 2014.
[123] Anonymous research participant #992-02, interview with Virginia Ferris, 3 April 2014.
[124] Anonymous research participant #992-02, interview with Virginia Ferris.

> "Linked data and RDF is not about filling out documents… It's really about creating a whole bunch of sentences or statements. And that's a barrier for archivists because many of them probably think that their job is to create finding aids… [Linked data] is not about creating a document. It's about asserting a whole myriad of facts. It's about asserting many, many, many, many facts on 'this thing has a title of – this thing has a creator of – this thing was written in – this thing has a subject term of – this thing is located at – ' It's a whole pile of little facts... You think of a form or a document as a story, right? Stories are made up of a whole pile of sentences. Instead of thinking about filling in a document to create a story, think about writing each sentence. That's what it is. It's about writing each sentence, and then those sentences can be combined later on by you or somebody else as a story."[125]

> "I think historically people have conceived of finding aids as almost like a published document that describes a collection in a very particular context, and it's very linear. I think there have been projects where people have tried to convert EAD to RDF, and so imagine every single component in a finding aid being a triple in RDF. I mean it's kind of hard to even think about. So an entity's relationship to its subseries or it has container "x", all of those relationships could be triples in theory and you could have this gigantic super-complicated looking pile of RDF triples that is really a finding aid. You could compile it all back together into a finding aid and it would look virtually the same to a researcher, but it would just require kind of a sea change in the way archivists think about what a record is."[126]

> "I don't know if there needs to be a shift in the way things are described because I do think the way archival description works is appropriate for what the community is trying to accomplish. I think if we made some strides with tooling, took some steps in terms of small things like I was just talking about with identifiers at the component level, and maybe kind of dissuade some of the archival community of some of their assumptions about inheritance of terms and access points up and down those hierarchies, and how that really plays out in practice. Little things like that, I think, would go a long way."[127]

Another issue cited by one participant was the question of how to ensure the reliability and trustworthiness of data sets, in addition to the lack of control over potential

---

[125] Eric Lease Morgan, interview with Virginia Ferris, 28 March 2014.
[126] Anonymous research participant #992-02, interview with Virginia Ferris, 3 April 2014.
[127] Corey Harper, interview with Virginia Ferris, 23 April 2014.

reuse in the increasing "mash-up of data"[128] and data reuse that exists on the web. The traditional finding aid preserves the core principles of provenance and contextual information for archivists through its hierarchical structure. Respondents indicated that some are concerned with the possibility that if practitioners begin to unpack collection descriptions into many smaller statements about items and relationships, archivists could run the risk of losing that important provenance and contextual information:

> "It's obviously very important to explore the whole area of provenance within linked data. That's one of the things that archivists worry about, because it's the idea of trusted sources of data. If you start effectively de-constructing and re-constructing data in all sorts of different ways, you know, where's the data coming from, is it trustworthy? If the Archives Hub starts providing this 'Linking Lives' interface to the public, are we saying we trust all the data that comes from these other data sources? Or is it the responsibility of the researcher? So I think there are really, really interesting questions. To be honest, whether it's linked data or otherwise, those questions aren't going to go away for archivists because of the nature of the web now. Even if you think you aren't doing linked data, where you're obviously doing that kind of mash-up of data, I still think we're all going in that direction either way."[129]

> "One of the principles of the web and linked data is something that's called the triple-A policy. Anybody can say anything, at any time. There's no centralized group that says this is what you're supposed to say and this is how you're supposed to say it. Anybody can assert a fact, even if it's false, and no one's going to get kicked out on a rail. Archivists aren't going to like that. It is entirely possible for someone to set up their RDF and associate it with something that's completely off the wall in their mind. People will say, 'That's blasphemy! That is so inappropriate! You are putting my content into the wrong context.' Well, maybe it is your content, but you're supposed to be sharing it with the world and it might put it into a different context."[130]

Several respondents commented on the challenge of finding quality data and the need for more consistent, complete, "linkable" data sets, and in particular the challenge of

[128] Jane Stevenson, interview with Virginia Ferris, 30 April 2014.
[129] Jane Stevenson, interview with Virginia Ferris.
[130] Eric Lease Morgan, interview with Virginia Ferris, 28 March 2014.

the work required to create data for unique archival entities and content that may not have existing data or URIs with which to link:

> "To really make linked data work, you have to use URIs, and they have to be URIs that take the shape of a URL that points to web servers. Another barrier is that when you have the content of archives, it's often times unique, and there are no well-established URIs for the people or organizations within a collection… It's entirely possible that a lot of content in archives is about people who haven't been written about in any place else but this particular archive. So it's hard to establish a URI for a particular person because the content is almost unique."[131]

> "It's a lot of work. I think that retrospectively it's almost too much work to think about retrospective linked data. I think going forward it's actually not very much work. I think the linking is definitely more work than the actually being able to link, because there are two steps: everything has to have a hard URI, and then once it has a hard URI you have to actually make the connection explicit, and that's linked data. So everything having to have hard URIs is really easy, but the actual choice you make in an explicit link, and then making that explicit link, is more difficult. We can do it automatically, but then we're linking everything and then it diminishes in value. Or we can do it intentionally and then the work is exponentially more. So it's kind of a rock and a hard place kind of situation: we either do it badly and fast, or we do it well and slow."[132]

> "When you actually try and create a front-end interface to the data, quite a lot of the time you end up going back to the same problem that you have with the transformation, which is the inconsistency of the data that you might find, the entities in your data, and matching them up to other data sets. The other challenge… is that we obviously wanted to link to other data sets, to pull in other data, but we didn't find a huge amount of useful linked data outputs that we could link to – bearing in mind we were working with mostly personal names. I think the quality of linked data varies quite a lot."[133]

> "Although there are lots of data sets out there, it doesn't necessarily mean that that data has been published in a way that's oriented to do things that are useful for you. Linking a range of different things and different people, what other people would want with publishing those perspectives in mind…. Just because you've got linked data doesn't mean that you can necessarily link that data. So that's certainly a challenge with linked data, just finding data to actually link."[134]

---

[131] Morgan, interview with Virginia Ferris.
[132] Kathy Wisser, interview with Virginia Ferris, 8 May 2014.
[133] Jane Stevenson, interview with Virginia Ferris, 30 April 2014.
[134] Adrian Stevenson, interview with Virginia Ferris, 30 April 2014.

Other participants referred to the difficulty of shifting practices in the archival community, both in technical and cultural senses. The technical challenges of legacy data and systems present an obstacle in shifting to new practices:

> "The biggest [challenges] are that there is a tremendous amount of legacy data in our world. There are a lot of legacy systems. There are a lot of vendors who are involved in this space. And there is a lot of institutional inertia that keeps us moving in a particular set of directions and makes it hard for us to change course when an opportunity like this presents itself."[135]

> "I think it's one of the tricky things that we'll continue having to serve, archives and EAD and EAC-CPF, and libraries, they've still got their heads in things like MARC, so we're having to still deal with the old as well as the new. But I think it might shift over. [Linked data] does have a lot of promise."[136]

In addition to this "institutional inertia," other participants commented on the challenge of introducing a new conceptual model and putting it into practice:

> "The weakest pace of [change] acts in the social dimension. Which is to say, people. And it really requires an agreed-upon semantics and everyone understanding it, and understanding it alike and applying it consistently. In my experience of human beings that's highly unlikely to happen except in very closed, tightly controlled environments where the potential for reasoning and inference machines is much more doable in the domain, and the control is shrunken. If we assume that we get sort of gradually over time an uptake of the new technologies and the new standards, there's the challenge of training people. Getting people to think about what they're doing in slightly different ways, and getting the developer community to work on tools that ameliorate the learning curve and technological challenge. There's a whole host of different kinds of challenges, some of them sociological, some of them technical. But ultimately the most daunting challenge is always the social."[137]

Several participants remarked upon the need for leading projects and examples of archives successfully implementing linked data practices, to give a concrete example of the uses and applications of linked data for archives:

---

[135] Corey Harper, interview with Virginia Ferris, 23 April 2014.
[136] Adrian Stevenson, interview with Virginia Ferris, 30 April 2014.
[137] Daniel Pitti, interview with Virginia Ferris, 6 May 2014.

> "Maybe you would need to have a big one doing something big, and then you could have a lot of people following. Because usually that's what's happening. In Europe you had the National Library of France who did something big, and then because you had the big one doing something really big, you had everybody following. People said, "The NLF is doing this, so let's do that as well." Maybe you would have to have a big consortium, a big archive in Europe or America doing something big, and then you would see more following."[138]

Making it work

Following a discussion of the perceived barriers, the participant was asked to pose recommendations and describe what they viewed as possible solutions to removing these barriers to adopting linked data in the archival community. Many respondents echoed that the creation of better tools for archivists to experiment with linked data would provide one step toward a solution. Most agreed that in order for linked data to progress within the archival community, there is a great need for user-friendly tools that will not disrupt existing workflows: "It needs to be easy. If they don't have to change their workflow too much, then I think it can be adopted in the next… 5 years."[139]

Several participants commented on exploring ways to leverage existing description and programs to build linked data:

> "I do advocate, though, the beginning process of creating RDF – I do advocate starting with EAD files or MARC records or even tab-delimited text files or relational databases, and then writing reports against them. Archivesspace, Archon, and Archivist's Toolkit are wonderful. People have lots of data in there, and it would be not extraordinarily difficult to generate RDF out of those databases that already exist. I think that would be a very good start. "[140]

---

[138] Valentine Charles, interview with Virginia Ferris, 10 April 2014.
[139] Eric Lease Morgan, interview with Virginia Ferris, 28 March 2014.
[140] Morgan, interview with Virginia Ferris.

Respondents also noted that examples of successful visualizations and interfaces for archival linked data would promote greater use and experimentation within the archival community:

> "I think what needs to happen is that some additional projects need to be demonstrated. Projects that are real and that don't over-sell…. Create an interesting, searchable, browseable interface to all of their collections… and enhance your searchable, browseable interface as well as your finding aid with content from outside, and improve their services. I'm going to then find ways to do the same thing with other archives across this continent. And through this process I would add pictures and sound and timelines and introductory paragraphs and create little bibliographies, and what would happen is that their collections would become a part of a greater whole and become more value-added. If that sort of application were created, then I think more archives would do it. Is it going to be adopted? Not right away. People need to see. Many people can't imagine big enough, so they need to see and have something demonstrated to them in order to go that route."[141]

Other participants recommended greater dialogue and open collaboration between different archives and cultural heritage institutions. Jane Stevenson commented that this approach has supported the work of the Archives Hub and Linking Lives project, and suggested that collaboration is critical to promoting common standards, greater interoperability, and openness of metadata and search functionality across domains:

> "My biggest recommendation would be that I think it's the sort of thing that needs to be done in collaboration… Our idea was that if we do this whole linking of archives in the UK community, then you've got an approach that is consistent for all the participating institutions that we potentially do linked data for. It will all be done in the same way. Whereas, if archivists go away and fiddle about outputting their own versions of RDF/XML or whatever, or JSON, whatever it might be, we'll end up having a whole lot of different uses of vocabularies and different URIs and things. So I think that's quite important, and I think that gives us an opportunity. It encourages us to collaborate, really. So while we've done this work, what we've tried to do is go out and talk to people and write blog posts that actually talk about it in great detail. So we're sharing it with the community, so people can understand exactly what we're doing and why."[142]

---

[141] Morgan, interview with Virginia Ferris.
[142] Jane Stevenson, interview with Virginia Ferris, 30 April 2014.

In addition to increased dialogue between archival and other cultural heritage institutions, many participants recommended more conversations between archivists and technicians, to support the likelihood of creating effective tools and to promote education and familiarity with linked data and archival description between these different communities. If archivists can engage with metadata specialists and developers, express to them the guiding values and needs of archival description, and learn more about linked data from them, it is more likely that archivists will be empowered to help create the tools that will allow them begin to use linked data:

> "I have a feeling maybe in terms of technical support that there might be less discussion between technical people and archivists. I know libraries have changed a lot and are employing much more technical people in their service to try to develop that. I'm not sure it's the case in archives. It is my feeling that maybe there is less, still not enough developers in archives or people who could bring these types of dialogue and give new ideas to more conservative archivists."[143]

> "At NYU I think we're very fortunate to have a subset of people within our special collections and archives communities who have a good sense of… 'computational thinking.' It's sort of the abilities and the approach that non-technical people need to have to the way they talk about systems and the way they build functional requirements so that they really are thinking through how things work. And I think that the more technologists see that, the more technologists then are willing to go the extra mile to understand the business practices and the systems analysis problems that are presented by these particular areas of practice, and I do feel like we're getting closer and closer to that. It's often two steps forward and one step back, but at least that's still progress."[144]

Participants often remarked on the need for more education and familiarity with linked data among archivists, to understand the potential uses and benefits:

> "It could help to make people in archives more familiar with the concept of linked data – what it is, what do you need for doing it, what are the tools available – because in fact in terms of data transformation a lot of mappings are already

[143] Valentine Charles, interview with Virginia Ferris, 10 April 2014.
[144] Corey Harper, interview with Virginia Ferris, 23 April 2014.

> available for them to convert their data. So maybe at first a bit of this education problem"
>
> "I think more projects that leverage linked data, a basic understanding of what it is beyond the sexiness of it, are needed because everyone's like, 'Ooh linked data, I've got to do that,' but folks that are saying that aren't really going below the surface of, 'I've heard a lot about it.' Just getting a more fundamental understanding of it and then allowing some space for creativity [is needed]…. We need to get comfortable with it first."
>
> "I think right at the top of the list [of solutions] is, of course, understanding [linked data] and developing an appreciation of it and what its potential is… So you have within an archive people who are interested and also have to be able to convince the people who control the resources that it's worth the trouble of spending time on it, that there are benefits to the archive and to the users of the archive that justify resources being devoted to this."[145]

Kathy Wisser further explained the need for education and experimentation to build on top of what is already existing and known, in order for new innovation to occur in the direction of linked data for archives, by citing the introduction of innovations ranging from the printing press to Encoded Archival Description:

> "One of the big things – and this is definitely true if you look at innovation – just think about when Gutenberg invented the printing press… When he invented the printing press, he printed a book that looked like a handwritten manuscript. And he did that because if it wasn't something that people were going to be comfortable with then it wouldn't have been accepted. That's why we have front matter in EAD, so we can have a title page. Because when we were actually producing finding aids in analog format, they had formal title pages. And we wanted to be able to say, 'No, this isn't something completely alien. It's a finding aid. It's just like the finding aid that you're used to. It's just in a structured format and it's online.' …We're not changing archival description, we're just changing the way that we communicate it. I think the same is true with all innovations, and explicitly linked data and RDF triples is just one of them."[146]

Participants also recommended that archivists engage in the creation of tools with an openness, curiosity, and willingness to experiment: "Figure out what's possible,

---

[145] Daniel Pitti, interview with Virginia Ferris, 6 May 2014.
[146] Kathy Wisser, interview with Virginia Ferris, 8 May 2014.

document where those gaps are, and then use that to define functional requirements around next-generation tools."[147] This openness and willingness to try new practices and ideas before they are fail-safe, many agreed, are critical to making progress toward joining other cultural heritage institutions in experimenting with linked data:

> "The problem is, I have a feeling that this domain [of archives] now has some delay compared to the libraries that are much more advanced, or even the museums now… People prefer to have everything perfect before jumping, when in fact libraries or museums are maybe more likely to make more concessions and like to do more little prototypes or to have more of an ongoing approach, rather than just waiting for the perfect thing and then going for it."[148]

> "I think we need, rather than individual archives coming up with ad hoc solutions, or unilateral solutions, the community needs to lean more and more on open source and shared solutions. Building small code sets that can be integrated in and used in various local systems to help do the things that they want to do."[149]

> "Experiment. Do some reading on these topics. Go listen to talks, watch videos of people giving talks. Think through what some of those possibilities are. Talk to the IT people at your shop, talk to the web developers at your shop. Figure out if they have any interest in this or if they've been tracking on some of these issues. Start thinking about the people and places and things. The resources in the semantic web or linked data sense that your archival descriptions are about – think about the relationships those may have with other institutions' holdings, with other library materials, museum materials and artifacts. And look at the extent to which having URIs there, and systems that would connect those URIs, would start creating a research environment where can more seamlessly find your things from somebody else's and vice versa. And try to prototype that. Learn a tiny bit of code and hack on something. Play with stuff. Experiment. See what's possible."[150]

Following earlier discussions of the challenge of inconsistent data for linking, several participants pointed to increased standardization of descriptive practices in

[147] Corey Harper, interview with Virginia Ferris, 23 April 2014.
[148] Valentine Charles, interview with Virginia Ferris, 10 April 2014.
[149] Daniel Pitti, interview with Virginia Ferris, 6 May 2014.
[150] Corey Harper, interview with Virginia Ferris, 23 April 2014.

archives and treatment of archival description as the creation of metadata as another step in the right direction:

> "There's definitely been a move to be more careful about treating archival description as data and less as just, you know, free text about collections. I think that as more people adopt linked data, people will be more careful about making sure that they're relating to things that have standard terms or URIs and not just typing things in."[151]

> "I really try to emphasize that in today's world, you want to create structured data because the more structured data is the more machine-processable it is… If the archivists who contribute the data [of a personal name], say, put in a name without life date – which they often do – then identifying that person in a machine-based way is pretty difficult. And they don't necessarily think of that, because they think, 'Well I've put the person name, and it's obvious who the person is in the context of the collection description.' But in the new world, it's about being able to de-couple these pieces of information and for them still to be meaningful."[152]

> "Linked data has the potential to simplify description. We still want to represent it at the multi-level, but at least we can start thinking of [the finding aid], of each piece of it as a record in itself, and not as a giant document that's a single record… I think people would be more careful about authoring finding aids, or would start considering every piece of data that they're entering as a record itself."[153]

Many participants pointed to the standard of Encoded Archival Context: Corporate Bodies, Persons, and Families (EAC-CPF) as a potential stepping-stone toward creating more linkable data within the archival community. In addition to EAC-CPF, the Social Networks and Archival Context (SNAC) Project, building national network of archival authority records for creators, was frequently cited as an example of how archives could begin to connect archival description with linked data:

---

[151] Anonymous research participant #992-02, interview with Virginia Ferris, 3 April 2014.
[152] Jane Stevenson, interview with Virginia Ferris, 30 April 2014.
[153] Anonymous research participant #992-02, interview with Virginia Ferris, 3 April 2014.

> "I think if the national EAC database [SNAC] pans out, that will be an easy win for archivists and linked data hopefully. I think people would adopt that, especially if tools that they're already using to write biographical notes about creators could churn out linked data or access it. Then people will use it and it could get adopted."[154]

Predictions and patterns for the future

Following discussions of possible solutions for removing barriers to implementing linked data in the archival community, participants gave their predictions for the future of archival description, with or without linked data. Many respondents spoke about the gradual process of change and adoption of new innovations, and proposed their theories on when, how, and if linked data may be incorporated into archival description in the future:

> "I've been proselytizing this stuff for a really long time and I realize I have to be very careful about framing it as some kind of panacea for all of the resource description problems we have in the cultural heritage sector. I don't know if this writ-large sort of grandiose idea of implementing linked data is even necessarily really a thing. I'm much more interested in incremental changes to descriptive practice that are taken with an awareness of what the web is and how the web works, in order to try to get small returns on investment."[155]

> "I don't think the archives community is in the best position to be the leading edge of linked data. I think people will adopt it as it becomes easy to adopt. There will be a few people who will undertake pilot projects and do really innovative things, but I don't know that profession-wide it will catch on quickly. I mean part of it is just that it's a huge endeavor. It has the potential to save time, but actually designing all the tools and systems to make things work is just a huge amount of work in itself."[156]

> "People always ask, 'Do you think linked data is worth doing?' And I really wanted to be able to answer that question. We started linked data 4 or 5 years ago, and in all honesty I'm still a bit on the fence about it, I hate to say. I can so see the benefits, and I think it's really exciting, the potential benefits. But at present it's

[154] Anonymous research participant #992-02, interview with Virginia Ferris.
[155] Corey Harper, interview with Virginia Ferris, 23 April 2014.
[156] Anonymous research participant #992-02, interview with Virginia Ferris, 3 April 2014.

> not trivial to do this kind of thing and to do it well. So I suppose my conclusion would be: If investment continues in linked data to the degree where you start getting tools and services that really help you to create good linked data, then I can see it really becoming a much bigger thing. But I think it's going to need that [investment]."[157]

Some respondents reiterated the importance of collaboration moving forward, across boundaries of different repositories, domains, and geography:

> "I think it's going to be all about collaboration and collective activity. Tiny little projects that are happening and that we've been doing forever... But I think that also accompanying those small, local initiatives, there are going to be national and international initiatives. The conversations are happening much more at a higher level of, 'What are the Australians doing?' That's the kind of conversations that are going on now. We're much more aware of companion institutions in other places, and we're interested in what they're doing."[158]

Building on previously discussed solutions, several participants commented on the increasing prevalence of "disaggregated archival description" as a sign that archivists are working toward linked data. Daniel Pitti remarked in particular on this movement toward creating archival description so that archivists can "separate out the data and maintain it independently and interrelate it," for reuse in a finding aid, a guide, or endless other possibilities for "creating other kinds of views into the data."[159] Pitti commented on the resilience and value of hierarchical description in archives, particularly as archivists contend with increasingly high volumes of born-digital records:

> "I don't think hierarchies are going to go away, in particular because there's a profound economy to them, and that economies will be lost if one simply dives into a vast network of interrelations. It potentially could be quite costly to compose and put together. So I don't think very economic approach that archivists take is going to go away simply because the quantitative problem that they face isn't going away, and in fact it's growing in magnitude. So that's going to stay, I think. But it does open it up to the possibility of creating new sets of

---

[157] Adrian Stevenson, interview with Virginia Ferris, 30 April 2014.
[158] Kathy Wisser, interview with Virginia Ferris, 8 May 2014.
[159] Daniel Pitti, interview with Virginia Ferris, 6 May 2014.

> relations and tying resources so that the descriptions of individual collections or *fonds* into a broader network of interrelations… In certain instances it doesn't introduce new work, it's working with what's already been done. But it demonstrates that that work also implicitly has within it the possibility of being drawn into a broader contextual environment. I often refer to this as being the 'great *fonds*.'"[160]

Pitti continued to remark on the evolution of communication technologies and archival description, providing his vision of the future of archival description:

> "If you look at the history of what's gone on in the libraries and museums and archives – particularly if you look at what's gone on in libraries, who tend to be ahead of the other two – there's been a shift. There is a trend and it's always the same. Communication technology opens up the possibility of furthering that drift toward reconfiguring, but the fundamentals of the way in which we describe remain the same. They're grounded in particular principles. But the way in which we represent that data for communication purposes, and what we can do with it after we've represented it that way. That I think is the drift of it all."[161]

Kathy Wisser expanded on her concern for creating "meaningful linked data," which would require the thoughtful linking of data that adds value to description rather than creates indiscriminate links between all resources that share a relationship. In a parallel example, Wisser described the new single search bar in the library website at Simmons College and observed that the "thousands and thousands and thousands" of records that show up in search results are not helpful for users, particularly those who perform known-item searches who are looking for a specific resource. Wisser explained:

> "In the same way, with linked data we're obliterating the concept of added value. We're obliterating the concept of expertise in favor of 'everybody wants everything they can get.' And that's not something that people do. I think it diminishes or dilutes the power of the tools that we have. So do I know which relationships are the important ones? No. I think that's a community decision. But I do think that we should be having a conversation, and that's the key: we need to have that conversation."[162]

---

[160] Pitti, interview with Virginia Ferris.
[161] Pitti, interview with Virginia Ferris.
[162] Kathy Wisser, interview with Virginia Ferris, 8 May 2014.

Wisser added that in her work with EAC-CPF she has focused on developing ways to "centralize information so that you can really target your biographical information in the finding aids on what is relevant to the archival collection that you have in front of you."[163] Wisser's explanation of meaningful linked data contributes an important discussion point for archivists to consider, by reiterating the importance of deciding what information is the most relevant and meaningful to describing a collection. In her explanation here, the need to create meaningful linked data reaffirms the need for description that is thoughtful and adds value to collections.

At the same time, Wisser's comments reflect an underlying understanding or concern shared among all of the participants for what users want. Jane Stevenson remarked on her observations of this realization in working with the Linking Lives project:

> "There's this idea that you're creating linked data and everything will be wonderful, but you then actually need the front end. You need a place where researchers can benefit from what you've done… Say you have a collection in archives, or several collections in archives that were created by George Bernard Shaw, and you're therefore just thinking in terms of the archives, but researchers often don't think like that. They more often think, 'I'm interested in George Bernard Shaw. I want to find information about this person.' So we wanted to bring archives within an environment where we were also showing works by the author or the person, or being able to link to Wikipedia to read more about the person, or whatever it might be."[164]

Participants frequently referred to their perception of user needs and search strategies that inform descriptive practices and interface designs. Among these statements about perceived use and user needs, Wisser commented on the need to confront the growing

---

[163] Wisser, interview with Virginia Ferris.
[164] Jane Stevenson, interview with Virginia Ferris, 30 April 2014.

misconception among users that, "if it isn't on the Internet, it doesn't exist."[165] Wisser expressed concern that this misconception is particularly damaging for archives, where the majority of materials are not digitized and on the web:

> **"**How do we argue against that, without just opening our doors and saying, 'See?' In that way I think that linked data also, the attention that linked data gets now, because it's hot and sexy, also can feed into that misconception."[166]

Following from Wisser's concern, while linked data presents the potential to more widely distribute description of archival materials online, it also necessitates greater advocacy and more effective expression of the diverse analog, digitized, and born-digital nature of archival collections to users. While linked data presents powerful opportunities, it also presents complicated new challenges in terms of representing the richness of collections and their relationships to other resources online.

[165] Kathy Wisser, interview with Virginia Ferris, 8 May 2014.
[166] Wisser, interview with Virginia Ferris.

CONCLUSIONS

While the sample size of research participants for this study was relatively small, responses provide a basic understanding of current attitudes toward and experiences with linked data in the archival community. These responses support several conclusions in regard to the current state of archival engagement with linked data and possible roads forward. The emphasis on user needs shared among research participants suggests a need for new user studies that can determine who these users are, what they actually want and need, and how they are searching, to provide data that will strengthen strategies moving forward for archivists looking to work with linked data. An assessment of user needs would provide the foundation for creating what Kathy Wisser termed "meaningful linked data."[167] Rather than linking everything that can be linked in archival description, archivists could determine the most meaningful, relevant, and available data to link and find the highest return for their investment of time and energy in linked data. Evidence of user needs could inform the way forward for archivists to link from the most authoritative, resourced, and trustworthy data sets, and to identify the entities and data sets that will provide the most meaningful information for users.

[167] Kathy Wisser, interview with Virginia Ferris, 8 May 2014.

Determining user needs will also be closely tied to the design of end-user interfaces that support diverse search strategies and provide multiple avenues to archival material, as outlined in Michael Whitelaw's proposal of the "generous interface," and through representational state transfer (REST) interfaces and other interface designs mentioned by respondents.[168] The concept of the generous interface supports "richer, more exploratory modes of information seeking" by creating more versatile, browseable interfaces guided by the following principles: "show first, don't ask," by volunteering "rich information that supports an audience's understanding of the collection;" providing "rich overviews" of collections that "help orient the user's exploration;" providing samples of content that "provide rich contextual cue and invite exploration;" providing context by displaying "the structure of relationships within the collection;" and sharing high quality primary content.[169] This proposed architecture for interfaces of archival collections is modeled in keeping with the "ethos of generosity," by "sharing liberally" information about collection materials and providing "seamless access" to that information.[170]

Participants' emphasis on the importance of greater education and familiarity with linked data in the archival community points to a need for integrating linked data into more widespread academic and professional dialogue. Incorporating semantic web and linked data principles and practices into graduate programs in archival studies could

---

[168] Michael Whitelaw, "Towards Generous Interfaces for Archival Collections," (paper presented at the International Council of Archivists Congress, Brisbane, Australia, 2012), accessed 1 July 2014, http://mtchl.net/towards-generous-interfaces-for-archival-collections/; Eric Lease Morgan, interview with Virginia Ferris, 28 March 2014; Anonymous research participant #992-02, interview with Virginia Ferris, 3 April 2014.
[169] Whitelaw, "Towards Generous Interfaces for Archival Collections," 7-8.
[170] Whitelaw, "Towards Generous Interfaces for Archival Collections," 7-8.

promote the use of and engagement with linked data in the next generations of archivists. Professional associations and conferences such as the Society of American Archivists and Code4Lib are also increasingly involved in supporting conversations about linked data and providing platforms for archivists and developers to collaborate. Greater dialogue and organized professional education around linked data, beginning in graduate school curricula, will help to remove the educational barrier to implementing linked data, as well as to empower more archivists to be able to engage with the technical and practical development of linked data practices for the archival community.

The recently published article "Toward an International Conceptual Model for Archival Description" by the International Council of Archivists, and discussed by research participant and co-author of the article Daniel Pitti, proposes a new framework for archival description that will support new communication technologies such as linked data. The article outlines the need for conceptual model of archival ontologies that brings together disparate international standards, and brings archival knowledge and data sets into the linked data sphere. The article acknowledges that the needs of the archival community to represent archival collections and single *fonds* places archival description somewhere between the "data-centric" model of database technologies and the "document-centric" model of markup technologies.[171] In acknowledgement of the continuing changes in these technologies and the unique needs of archival description, the ICA proposal advances:

---

[171] Gretchen Gueguen, et al, "Toward an International Conceptual Model for Archival Description: A Preliminary Report from the International Council on Archives' Experts Group on Archival Description," *The American Archivist*, Vol. 76, No. 2 (Fall/Winter 2013): 572.

> "The broader and more persistent objective of separating components of description, so as to have a more economical and flexible method of creating and maintaining the data that comprises description. Once separated, the components can be exploited to create a wide variety of discovery and access systems that multiply the use perspectives, all while maintaining the essential archival *respect des fonds*."[172]

The leading principles of ICA's proposal to create new international standards for archival description will allow archives to engage more with semantic technologies of linked data, by providing a guide for archivists to create structured data about their collections that can be more easily "computationally exploited to reveal new knowledge, new patterns, and thus an information universe that is more than the sum of its parts," and that will transcend the boundaries of separate institutions and domains by interconnecting "data that are created and maintained in disparate contexts, and thus overcome the separation between different descriptive systems and the things they describe."[173] The eventual goal of creating archival description that can more easily engage with linked data and semantic technologies will allow archives to become a part of what Charles Jewett termed the "universal catalog of cultural heritage," by de-siloing separate domains in order to "explore interconnecting disparate description systems and realize integrated access to cultural heritage," and that will "support multiple paths into archival resources."[174] The ICA has formed an Experts Group on Archival Description (EGAD)

---

[172] Guegen, et al, "Toward an International Conceptual Model for Archival Description: A Preliminary Report from the International Council on Archives' Experts Group on Archival Description," 570.
[173] Guegen, et al, "Toward an International Conceptual Model for Archival Description: A Preliminary Report from the International Council on Archives' Experts Group on Archival Description," 573.
[174] Charles C. Jewett, *On the Construction of Catalogues of Libraries and Their Publication by Means of Separate, Stereotyped Titles* (Washington, D.C.: Smithsonian Institution, 1853); Guegen, et al, "Toward an International Conceptual Model for

that will develop this international standard, to be released in 2016 as a formal document and represented in Web Ontology Language (OWL) that will "facilitate its use immediately in a variety of semantic technologies… exposing archival description using LOD [Linked Open Data] techniques."[175]

As explained by Eric Lease Morgan and echoed by other research participants, the promise of linked data as currently perceived by practitioners, and reflected in the developments of new interfaces and standards for description, is guided by a desire to "tell a much richer, more thorough story" about archival collections and place those stories on a more prominent stage on the open web with other cultural heritage materials.[176] Despite the many challenges currently standing in the way of widespread use and acceptance of semantic technologies in the archival community, there appears to be growing awareness of the potential for linked data to allow archivists to move beyond "showing what we have," to showing how their collections speak to other collections and resources beyond the walls of an individual repository.

Further research and case studies examining applications of linked data in the archival community will build a common conceptual understanding and trust of it among archival practitioners, encouraging greater use and providing clear reference models for using linked data as a reliable and beneficial practice. Greater trust and understanding of the practice among the archival community should contribute to more widespread

---

Archival Description: A Preliminary Report from the International Council on Archives' Experts Group on Archival Description," 573 and 578.

[175] Guegen, et al, "Toward an International Conceptual Model for Archival Description: A Preliminary Report from the International Council on Archives' Experts Group on Archival Description," 579.

[176] Eric Lease Morgan, interview with Virginia Ferris, 28 March 2014.

adoption and allocation of resources to fund further development of linked data for archival description. The potential for linked data, as stated by participants in this study, centers on its ability to support and link archival description, increasing the visibility, discovery, and use of archival materials. These effects could benefit archival advocacy efforts by exposing archival collections, engaging broader populations of researchers, and demonstrating the immense value and interconnectedness of archives with the wider universe of information resources in the semantic web.

BIBLIOGRAPHY

Alemu, Getaneh, Brett Stevens, Penny Ross, and Jane Chandler. "Linked Data for Libraries: Benefits of a Conceptual Shift from Library-Specific Record Structures to RDF-based Data Models." In *New Library World,* Vol. 113, No. 11-12 (2012).

Allemang, D., and J. Hendler. *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*. Boston, Massachusetts: Morgan Kaufman Publishers, 2008.

Anonymous research participant #992-02. Interview with Virginia Ferris. 3 April 2014.

Archives Hub. "Home." Archives Hub: At the Center of Great Research. Accessed 1 July 2014. http://archiveshub.ac.uk/.

Archives Hub. "Introducing the Archives Hub." Archives Hub: At the Center of Great Research. Accessed 1 July 2014. http://archiveshub.ac.uk/introduction/.

Archives Hub. "Linking Lives Evaluation Report." Archives Hub: At the Center of Great Research. Accessed 1 July 2014. http://archiveshub.ac.uk/linkinglives/.

Australian Policy Online. "Linked Data Project." Accessed 7 July 2014. http://apo.org.au/content/linked-data-project.

Berners-Lee, Tim. "The Semantic Web." In *Scientific American*, 284, (2001): 34-43.

Boles, Frank. 64th Presidential Address of the Society of American Archivists. Presented at the Society of American Archivists Annual Meeting, Austin, Texas, 14 August 2009.

British National Bibliography. "Welcome to http://bnb.data.bl.uk/." British National Bibliography as Linked Open Data. Accessed 1 July 2014. http://bnb.data.bl.uk/.

Brown, Jesse. "More Than Just a Name: Archival Authority Control, Creator Description, and the Development of Encoded Archival Context (EAC)." Master's paper, School of Information and Library Science, University of North Carolina at Chapel Hill, 2006.

Byrne, Gillian, and Lisa Goddard. "The Strongest Link: Libraries and Linked Data." *DLib Magazine,* Vol. 16, No. 11-12 (2012). Accessed 1 July 2014. http://www.dlib.org/dlib/november10/byrne/11byrne.html.

Charles, Valentine. Interview with Virginia Ferris. 10 April 2014.

Decipher Project. "Introduction." Decipher Project. Accessed 1 July 2014. http://decipher-research.eu/introduction.

Digital Public Library of America. "Welcome." Digital Public Library of America. Accessed 1 July 2014. http://dp.la/.

Digital Repository of Ireland. "Linked Logainm: Placenames Database." Digital Repository of Ireland. Accessed 1 July 2014. http://www.logainm.ie/ga/.

Digital Repository of Ireland. "LocationLODer." Digital Repository of Ireland. Accessed 1 July 2014. http://apps.dri.ie/locationLODer/.

Digital Repository of Ireland. "Projects." Digital Repository of Ireland. Accessed 1 July 2014. http://dri.ie/projects.

Dow, Elizabeth H. "Encoded Archival Description as a Halfway Technology." *Journal of Archival Organization*, 7:3, (2009): 108-115.

Encoded Archival Context Working Group of the Society of American Archivists and the Staatsbibliothek zu Berlin. "EAC-CPF Tag Library." Version 2010 Draft. Accessed 1 July 2014. http://www3.iath.virginia.edu/eac/cpf/tagLibrary/cpfTagLibrary.html.

Engels, Robert H.P., and Jon Roar Tonneson. "Case Study: A Digital Music Archive (DMA) for the Norwegian National Broadcaster (NRK) Using Semantic Web Techniques." *World Wide Web Consortium: Semantic Web Use Cases and Studies* (2007). Accessed 1 July 2014. http://www.w3.org/2001/sw/sweo/public/UseCases/NRK/.

Europeana. "Europeana Data Model (EDM) Documentation." Europeana Professional. Accessed 1 July 2014. http://pro.europeana.eu/edm-documentation.

Europeana. "Portal Homepage." Europeana. Accessed 1 July 2014. http://europeana.eu/portal/.

Europeana. "Task Force on Hierarchical Objects." Europeana Professional. Accessed 1 July 2014. http://pro.europeana.eu/web/network/europeana-tech/-/wiki/Main/Taskforce+on+hierarchical+objects.

Europeana. "Thought Lab." Europeana. Accessed 1 July 2014. http://pro.europeana.eu/web/guest/linked-open-data.

Europeana Task Force on Hierarchical Objects. "Recommendations for the Representation of Hierarchical Objects in Europeana." Europeana Professional. Accessed 1 July 2014. http://pro.europeana.eu/documents/468623/4a6eb2ec-4cc6-48b1-8824-92a1e564a279.

Free Your Metadata. "Access." Ghent University, iMinds and MASTIC, Universite de Bruxelles. 2011-2014. Accessed 1 July 2014. http://freeyourmetadata.org/sustainable-access.

Free Your Metadata. "Cleanup." Ghent University, iMinds and MASTIC, Universite de Bruxelles. 2011-2014. Accessed 1 July 2014. http://freeyourmetadata.org/cleanup.

Free Your Metadata. "Extraction." Ghent University, iMinds and MASTIC, Universite de Bruxelles. 2011-2014. Accessed 1 July 2014. http://freeyourmetadata.org/named-entity-extraction.

Free Your Metadata. "Reconcile." Ghent University, iMinds and MASTIC, Universite de Bruxelles. 2011-2014. Accessed 1 July 2014: http://freeyourmetadata.org/reconciliation.

Greenberg, Jane, and Eva Mendez. "Linked Open Data for Open Vocabularies and HIVE's Global Framework." *El Profesional de la Informacion,* Vol. 21, No. 3 (2012).

Gueguen, Gretchen, Vitor Moel Marques da Fonseca, Daniel V. Pitti, and Claire Sibille-de Grimoulard. "Toward an International Conceptual Model for Archival Description: A Preliminary Report from the International Council on Archives' Experts Group on Archival Description." *The American Archivist,* Vol. 76, No. 2 (Fall/Winter 2013).

Harper, Corey. Interview with Virginia Ferris. 23 April 2014.

Hensen, Steve. "Squaring the Circle: The Reformation of Archival Description in AACR2." *Library Trends* 36 (Winter 1988).

Isaac, Antoine. "Case Study: Enriching and Sharing Cultural Heritage Data in Europeana." *World Wide Web Consortium: Semantic Web Use Cases and Studies.* June 2012. Accessed 1 July 2014. http://www.w3.org/2001/sw/sweo/public/UseCases/Europeana/.

Invisible Australians: Living Under the White Australia Policy. "Home." Invisible Australians Project, Kate Bagnall and Tim Sherratt. Accessed 6 July 2014. http://invisibleaustralians.org/.

Jewett, Charles C. *On the Construction of Catalogues of Libraries and Their Publication by Means of Separate, Stereotyped Titles*. Washington, D.C.: Smithsonian Institution, 1853.

Kilfeather, Eoin. "Deciphering the Story of the Museum." *ERCIM News*, July 2011. Accessed 1 July 2014. http://ercim-news.ercim.eu/en86/special/deciphering-the-story-of-the-museum.

LiAM Project: Linked Archival Metadata. "Prospectus for Linked Archival Metadata: A Guidebook." *LiAM: Linked Archival Metadata Blog,* 1 March 2013. Accessed 1 July 2014. http://sites.tufts.edu/liam/deliverables/prospectus-for-linked-archival-metadata-a-guidebook/.

Library of Congress Linked Data Service, Authorities and Vocabularies. "About." Library of Congress. Accessed 1 July 2014. http://id.loc.gov/about/.

Library of Congress Linked Data Service, Authorities and Vocabularies. "Home." Library of Congress. Accessed 1 July 2014. http://id.loc.gov/.

LOCAH Linked Archives Hub. "About." LOCAH Linked Archives Hub. Accessed 1 July 2014. http://data.archiveshub.ac.uk/about.html.

MacNeil, Heather. "Metadata Strategies and Archival Description: Comparing Apples to Oranges." *Archivaria* 39 (April 1995).

Meehan, Jennifer. "Making the Leap from Parts to Whole: Evidence and Inference in Archival Arrangement and Description." *American Archivist* 72 (Spring/Summer 2009): 72-90.

Millar, Laura. *Archives: Principles and Practices*. New York, NY: Neal-Schuman Publishers, Inc., 2010.

Mitchell, Erik T. "Three Case Studies in Linked Open Data." *Library Technology Report, Vol. 49,* No. 5 (2013): 26-43.

Morgan, Eric Lease. Interview with Virginia Ferris. 28 March 2014.

Morgan, Eric Lease, and the LiAM: Linked Archival Metadata Project. *Linked Archival Metadata: A Guidebook*. Version 0.99, 23 April 2014. Accessed 1 July 2014. http://infomotions.com/sandbox/liam/tmp/guidebook.pdf.

Morrissey Sheila M. "More What You'd Call 'Guidelines' than Actual Rules: Variation in the Use of Standards." *Journal of Electronic Publishing*, Vol. 14, Issue 1

(Summer 2011). Accessed 1 July 2014. http://dx.doi.org/10.3998/3336451.0014.104.

Mulholland, Paul. "Curate: An Ontology for Describing Museum Narratives." *Decipher Project*, 1 November 2013. Accessed 1 July 2014. http://decipher.open.ac.uk/curate/introduction.

National Archives of Australia. "AGLS." National Archives of Australia AGLS Metadata Standards. Accessed 7 July 2014. http://www.naa.gov.au/records-management/agency/create-capture-describe/describe/agls.aspx.

OCLC. "Virtual International Authority File (VIAF)." OCLC Inc. Accessed 1 July 2014. http://www.oclc.org/viaf.en.html.

Oomen, Johan, Lotte-Belice Baltussen, and Marieke Van Erp. "Sharing Cultural Heritage the Linked Open Data Way: Why You Should Sign Up." Paper presented at *Museums and the Web*, San Diego, CA, 2012. Accessed 1 July 2014. http://www.museumsandtheweb.com/mw2012/papers/sharing_cultural_heritage_the_linked_open_data.

Pitti, Daniel. Interview with Virginia Ferris. 6 May 2014.

Research Data Australia. "Linked Data Policy Hub – connected resources for social research." Research Data Australia, Activity. Accessed 6 July 2014. http://researchdata.ands.org.au/linked-data-policy-hub-connected-resources-for-social-research.

Ruddock, Bethan. "Linked Data and the LOCAH Project." *Business Information Review*, Vol. 28, No. 2 (2011): 105-111.

Scott, Dan. "Structured Data NOW: Seeding schema.org in library systems." Abstract in *Code4Lib* 2014 conference program. Accessed 1 July 2014. http://code4lib.org/conference/2014/scott.

Scott, Dan. "Structured Data NOW: Seeding schema.org in library systems." Slides presented at *Code4Lib*, Raleigh, NC, 2014. Accessed 1 July 2014. http://stuff.coffeecode.net/2014/structured_data_now/#/8.

Tim Sherratt. "It's all about the stuff: Collections, interfaces, power, and people." Invisible Australians: Living Under the White Australian Policy. Accessed 6 July 2014. http://invisibleaustralians.org/blog/2011/12/it%E2%80%99s-all-about-the-stuff-collections-interfaces-power-and-people/.

SNAC: Social Network and Archival Context Project. "SNAC Home." SNAC: Social Network and Archival Context Project. Accessed 1 July 2014. http://socialarchive.iath.virginia.edu/.

Society of American Archivists. *Describing Archives: A Content Standard*. Chicago, IL: Society of American Archivists, 2013.

Society of American Archivists. "Encoded Context for Archives: Corporate Bodies, Person, and Families (EAC-CPF)." Society of American Archivists, Standards. Accessed 1 July 2014. http://www2.archivists.org/groups/technical-subcommittee-on-eac-cpf/encoded-archival-context-corporate-bodies-persons-and-families-eac-cpf.

Stevenson, Adrian. Interview with Virginia Ferris. 30 April 2014.

Stevenson, Adrian. "LOCAH Project: Wider Benefits to Sector and Achievements for Host Institution." *LOCAH Project* blog, 2010. Accessed 1 July 2014. http://archiveshub.ac.uk/locah/2010/08/locah-project-wider-benefits-to-sector-achievements-for-host-institution/.

Stevenson, Jane. Interview with Virginia Ferris. 30 April 2014.

Stevenson, Jane. "'What Happens If I Click on This?': Experiences of the Archives Hub." *Ariadne*, no. 57 (October 2008). Accessed 1 July 2014. http://www.ariadne.ac.uk/issue57/stevenson.

Voss, Jon. "LODLAM State of Affairs." *Information Standards Quarterly*, Vol. 24, No. 2/3 (Spring/Summer 2012): 41-42.

Voss, Jon. "Radically Open Cultural Heritage Data on the Web." Presented at Museums and the Web, San Diego, CA, 11-14 April 2012. Accessed 1 July 2014. http://www.museumsandtheweb.com/mw2012/papers/radically_open_cultural_heritage_data_on_the_w.

W3C Schema Bib Extend Community Group. "About." W3C Community and Business Groups. Accessed 1 July 2014. http://www.w3.org/community/schemabibex/.

Whitelaw, Michael. "Towards Generous Interfaces for Archival Collections." Presented at the International Council of Archivists Congress, Brisbane, Australia, 2012. Accessed 1 July 2014. http://mtchl.net/towards-generous-interfaces-for-archival-collections/.

Wisser, Kathy. Interview with Virginia Ferris. 8 May 2014.

Wolff, Annika, Paul Mulholland, and Trevor Collins. "Storyspace: A Story-Driven Approach for Creating Museum Narratives." Milton Keynes, United Kingdom: Knowledge Media Institute, the Open University, 2012. Accessed 1 July 2014. http://people.kmi.open.ac.uk/paulm/hypertext-2012.pdf.

Zengenene, Dydimus. "Conference Report: Global Interoperability and Linked Data in Libraries." *New Library World*, Vol. 114, No. 1-2 (2013): 84-87.

Appendix A

Interview Questions

1. Where do you currently work and what do you do there?
2. What is linked data? Please define it in your own words.
3. What is your background with linked data? When did you first learn about it? How have you used it?
4. Are you working with linked data at your current institution? [*If "Yes," ask the following.*]
    a. In what ways are you using it?
    b. What were the greatest practical or conceptual challenges **t**hat you encountered?
    c. What do you think might have made linked data easier to use?
    d. What do you perceive to be/did you discover to be the greatest benefits for the work that you do? Potential benefits?
5. What do you see as the greatest potential value and benefits of linked data in archives?
6. What are the greatest challenges or barriers to using linked data in archives?
7. What might remove some of these barriers?
8. Do you think linked data could be widely adopted by archivists?
9. Do you think linked data could be used to expose "hidden" archival collections online? Why or why not?
10. How do you think linked data might impact researchers' discovery and use of archives?
11. How do you think linked data might impact the practice of archival description moving forward?
12. What examples have you seen of successful projects using linked data?
13. Based on your experience or observations, what recommendations would you give to other institution considering linked data as a practice?

Appendix B

Interview Subject Heading Codes

1. Background
    - A. Defining linked data
    - B. Using/learning about linked data
    - C. Current uses of linked data in their work
2. Potential value
    - A. Enriching description and representation of collections
        - I. Linking with diverse data sources (de-siloing)
        - II. Leveraging existing description
        - III. Increasing greater contextualization
    - B. Increasing discovery and use
        - I. Exposing “hidden” collections
    - C. Improving consistency, standardization, quality of archival metadata
    - D. Facilitating reuse of data
    - E. Saving time with more efficient description workflows
        - I. Sharing/building on existing description
3. Barriers
    - A. Technical complexity
    - B. Need for tools
    - C. Need for concrete examples, leaders
    - D. Traditional archival description
        - I. Hierarchical structure
        - II. Single record, document focus
    - E. How to preserve/represent provenance and context
        - I. How to ensure trustworthiness of data
        - II. Archival principles retained in finding aid
    - F. Legacy data and systems, cultural/social challenges of major change
        - I. Need for conceptual model/new data model/standards
    - G. Need to treat description as creation of metadata, not just document markup
    - H. Inconsistent quality of archival data/URIs/data sets to link with
        - I. Requires a lot of work up front to clean up or find/create good data
4. Making it work

A. Leveraging existing tools/database models – Archivist's Toolkit, Archivesspace
B. Creating/providing more concrete examples – visualizations, interfaces
C. Greater dialogue with technicians
    I. Engaging in creation of tools
D. Collaboration within archives/cultural heritage communities
E. More education and familiarity with the concepts
F. Openness, curiosity, willingness to play, experiment, interrogate
G. Increased standardization
    I. Treating description as creation of metadata
H. List of current examples of interesting/successful projects

5. Conclusions

A. Predictions re: future of archival description and linked data
B. Other thoughts
    I. Meaningful linked data
    II. Thinking about role of archivist and researcher
    III. Archival description as essence of linked data (relationships)